ABSTRACT

XU, ZEKUN. Modern Methods of Hidden Markov Models and Partially Observable Markov Decision Processes in Biostatistics. (Under the direction of Eric B. Laber and Ana-Maria Staicu.)

Hidden Markov models and partially observable Markov decision processes are widely used to model system dynamics driven by latent states. Those latent states have specific practical interpretations in different real world problems, like the activity states in wearable computing and the mood states in bipolar disorder patients. Since the states are not directly observable, modern techniques are needed in order to estimate the state transitions and conditional state distributions. This thesis aims to develop novel computational and inferential methodologies to address scientific questions in biostatistics applications using hidden Markov model and partially observable Markov decision processes.

In the first part, the motivating wearable computing project comes from a randomized clinical study to evaluate the effectiveness of a nonsteroidal anti-inflammatory drug in owned, pet adult domestic cats. Accelerometer is used for collecting high-frequency activity data on cats and subsequently to study the effects of treatment on pain related activity patterns. However, activity patterns and the effect of treatment on these patterns can vary widely across cats; furthermore, the effects of treatment may manifest in higher or lower activity counts or in subtler ways like changes in the frequency of certain types of activities. We use a zero inflated Poisson hidden semi-Markov model to characterize activity patterns and subsequently derive estimators of the treatment effect in terms of changes in activity levels or frequency of activity type. We demonstrate the application of our model using data from a naturally occurring feline osteoarthritis-associated pain model.

In the second part, we continue with modeling high-frequency human activity data collected from wearable devices in the National Health and Nutrition Examination Survey. However, modeling such data is challenging as they are high-dimensional, heterogeneous, and subject to informative missingness, e.g., zero readings when the device is removed by the participant. We propose a flexible and extensible continuous-time hidden Markov model to extract meaningful activity patterns from human accelerometer data. To facilitate estimation with massive data we derive an efficient learning algorithm that exploits the hierarchical structure of the parameters indexing the proposed model. We also propose a bootstrap procedure for interval estimation.

In the last part, we shift the attention to the estimation of optimal treatment regime in the observational pathway of the Systematic Treatment Enhancement Program for Bipolar

Disorder. In this observational longitudinal study, no two patients have the same treatment history even after coarsening clinic visits to a weekly time-scale. Estimation of an optimal treatment regime using such data is challenging as one cannot naively pool together patients with the same treatment history; nor is it possible to apply backwards induction over the decision points. Thus, additional structure is needed to effectively pool information across patients and within a patient over time. Current scientific theory for many chronic mental illnesses maintains that a patient's disease status can be conceptualized as transitioning among a small number discrete states. We use this theory to inform the construction of a partially observable Markov decision process model of patient health trajectories wherein observed health outcomes are dictated by a patient's latent health state. Using this model, we derive an estimator of an optimal treatment regime under two common paradigms for quantifying long-term patient health. The finite sample performance of the proposed estimator is demonstrated through a series of simulation experiments and application to the observational pathway of the STEP-BD study. We find that the proposed method provides high-quality estimates of an optimal treatment strategy in settings where existing approaches cannot be applied without *ad hoc* modifications.

Modern Methods of Hidden Markov Models and Partially Observable Markov Decision Processes in Biostatistics

by
Zekun Xu

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2020

APPROVED BY:

Marie Davidian	B. Duncan X. Lascelles
Eric B. Laber Co-chair of Advisory Committee	Ana-Maria Staicu Co-chair of Advisory Committee

BIOGRAPHY

Zekun Xu was born in Shanghai, China. He received his Bachelor's degree in Economics from Fudan University in 2012. He continued his study at Carnegie Mellon University and obtained a Master's degree in Statistical Practice. After that, he worked as Biostatistician at Cleveland Clinic and INC Research. In 2015, he entered the PhD in Statistics program at North Carolina State University in pursuit of a Doctoral degree in Statistics. His research interests include hidden Markov models, reinforcement learning, and dynamic treatment regimes.

ACKNOWLEDGEMENTS

I would like to thank my thesis advisors Dr. Eric Laber and Dr. Ana-Maria Staicu for their patient guidance and kind support throughout my PhD study. I am grateful to all their feedbacks and suggestions regarding my research project both during our in-person and remote meetings. Not only do they broaden my statistical expertise but they also show me the qualities of a good researcher through numerous thought-provoking discussions we have had. I am also thankful to them for their lab meetings where I have the opportunity to communicate with other outstanding fellow students about our research.

I would like to thank my committee members Dr. Marie Davidian and Dr. Duncan Lascelles for their valuable comments on my thesis. My first important research project was through collaboration with Duncan, who opened my world about how statistical inference can help with veterinary studies. In addition to being one of my committee members, Marie also taught me Statistical Methods for Analysis With Missing Data and Dynamic Treatment Regime, both of which have tremendous impact on improving my statistical thinking and research methodology.

I would like to thank all the professors, faculty, and friends from the Statistics Department at NC State University for providing their help and support during the past five years. I would like to thank all colleagues from the IoT Department at SAS Institute, especially Dev Kakde, Arin Chaudhuri, and Gul Ege, who I worked closely with and delivered many quality results together.

Finally, I would like to express my appreciation to my parents for their endless love. I would like to express my deepest love and gratitude to my girlfriend, Ye Liu, for all the wonderful moments that we share.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

1.1 Overview

The class of Markov processes and its variants have been widely used to model dynamic systems in a variety of applications [see Dra67, for an overview on Markov process]. The simplest form of a Markov process is a Markov chain, which is a stochastic process $\{S^t\}_{t\geq 1}$ with a finite state space $S^t \in \{1,\ldots,M\}$ and satisfies the Markov property for all $t > 1$:

$$P(S^t = s^t | S^{t-1} = s^{t-1},\ldots,S^1 = s^1) = P(S^t = s^t | S^{t-1} = s^{t-1}).$$

In many real world applications, the states are not directly observable; instead, what we often observe is some noisy realization of the latent states. For example, when analyzing the physical activity data collected by wearable monitors, we only observe the activity intensity for each subject, whereas the actual type of activity (e.g. walking, running, etc.) corresponding to the activity intensity is unknown. Also in the example of estimating an optimal rule for treating bipolar disorder patients, we do not observe the true mood states but only the patients' answers to an inventory of mood-related questions. These applications motivate the development of hidden Markov models (HMM) and partially observable Markov decision processes (POMDP), which are useful frameworks in the literature to model dynamic systems driven by latent states. In this thesis, we will focus on modern methods of hidden Markov models and partially observable Markov decision processes that are motivated by various biostatistics applications.

1.1.1 Hidden Markov models

HMM was first introduced in the 1960s [BP66; Bau70]. Since then, there have been many extensions to the original HMM framework, including hidden semi-Markov model [Fer80; RM85; Lev86; BB06; Yu15], factorial hidden Markov model [GJ96; Gae09], Markov-switching model [Ham89; Ham96; So98], continuous-time hidden Markov model, [Bur03; Liu15] etc.. There have been many successful applications of hidden Markov model to solve real world problems in the field of speech recognition, computational biology, financial economics, and activity modeling [see, for example, Rab89; Ham89; Cap09; LZ11; LB17]. In particular, HMM is a common approach in the analysis of wearable sensor data [He07; ZS09; MS10; Nic11; LC11; RC14]. In this framework, the human activity is divided into different latent behavioral states. The observed activity count is dependent on the unobserved latent state. In the context of treatment evaluation, HMMs have also been used to model latent health states and subsequently conduct inference for activity patterns in terms of transitions among these states and activity patterns within each state [Sco05; Alt07; Shi10].

We will use the activity intensity data measured by wearable accelerometer as a concrete example to illustrate the framework for statistical inference in HMM. This type of data are often right-skewed integers, with an excessive proportion of zero readings, which motivates the following zero-inflated Poisson hidden Markov model. Relevant notations will be defined and theoretical properties will be discussed. We assume that the observed data are of the form $\left\{\left(\mathbf{X}_i^1, Y_i^1, \ldots, \mathbf{X}_i^T, Y_i^T\right)\right\}_{i=1}^n$ and comprise n independent replicates of the trajectory $\left(\mathbf{X}^1, Y^1, \ldots, \mathbf{X}^T, Y^T\right)$, where: $T \in \mathbb{N}_+ = \{1, 2, \ldots\}$ is a fixed time horizon; $\mathbf{X}^t$ denotes the subject covariates at time $t = 1, \ldots, T$; and $Y^t \in \mathbb{N} = \{0\} \bigcup \mathbb{N}_+$ denotes the integer accelerometer reading at time $t = 1, \ldots, T$. Let $S_i^t \in \{1, ..., M\}$ be the latent state at time t for subject i, where M is a fixed integer. This latent state trajectory evolves according to a Markov model indexed by: (i) an initial state distribution $\delta_i(m) \triangleq P(S_i^1 = m)$ for $m = 1, ..., M$; (ii) the transition probabilities $\mathbf{Q}_i(m, \ell; \mathbf{x}^t) \triangleq P(S_i^{t+1} = \ell | S_i^t = m, \mathbf{X}_i^t = \mathbf{x}^t)$ for $m, \ell = 1, ..., M$. Assume a multinomial logistic regression model for the transition probabilities to capture the potential treatment effect on transition probabilities so that

$$Q_i(m, \ell; \mathbf{x}^t) = \frac{\exp\left(\mathbf{d}'_{m,\ell,i} \mathbf{x}^t\right)}{\sum_{k=1}^M \exp\left(\mathbf{d}'_{m,k,i} \mathbf{x}^t\right),}$$

where $\mathbf{d}_{m,k,i}$ are unknown coefficient vectors for $m,k=1,\ldots,M$, $i=1,\ldots,n$. Further assume that $Y_i^t \perp (\mathbf{X}_i^1, S_i^1, \ldots, \mathbf{X}_i^{t-1}, S_i^{t-1}) \big| (\mathbf{X}_i^t, S_i^t)$ and that the accelerometer activity readings are distributed as

$$\begin{aligned}
&P\left(Y_i^t = y \middle| S_i^t = 1, \mathbf{X}_i^t = \mathbf{x}^t\right) \\
=&p_i^t(\mathbf{x}^t) 1_{y^t=0} + \left\{1 - p_i^t(\mathbf{x}^t)\right\} \frac{\left\{\lambda_i^t(m,\mathbf{x}^t)\right\}^y \exp\left\{\lambda_i^t(m,\mathbf{x}^t)\right\}}{y!},
\end{aligned}$$

when the subject is in the first latent state, and

$$P\left(Y_i^t = y \middle| S_i^t = m, \mathbf{X}_i^t = \mathbf{x}^t\right) = \frac{\left\{\lambda_i^t(m,\mathbf{x}^t)\right\}^y \exp\left\{\lambda_i^t(m,\mathbf{x}^t)\right\}}{y!},$$

when the subject is in latent states $m=2,3,\ldots,M$. Thus, accelerometer readings are modeled a zero-inflated Poisson with intensity functions $\lambda_i^t : \mathscr{S} \times \mathbb{R}^q \times \mathbb{R}^p \to \mathbb{R}_+$ and weight functions $p_i^t : \mathscr{S} \times \mathbb{R}^q \times \mathbb{R}^p \to [0,1]$ for $i=1,\ldots,n$, $t=1,\ldots,T$. Furthermore, for each subject i, time t, and latent state m, assume that these functions are of the form

$$\begin{aligned}
\log\left\{\frac{p_i^t(\mathbf{x}^t)}{1-p_i^t(\mathbf{z}^t,\mathbf{x}^t)}\right\} &= b_{0,0,i} + \mathbf{b}'_{1,0,i}\mathbf{x}^t \\
\log\left\{\lambda_i^t(m,\mathbf{x}^t)\right\} &= b_{0,m,i} + \mathbf{b}'_{1,m,i}\mathbf{x}^t,
\end{aligned}$$

where $b_{0,k,i}, \mathbf{b}_{1,k,i}$, $k=0,1,\ldots,M$ are unknown coefficient vectors.

In order to compute the likelihood of the hidden Markov model, we need to utilize the forward-backward recursion [Rab89] to define the following quantities. For subject $i=1,...,n$, define the forward variables for $t=1,...,T$, and $m=1,...,M$,

$$\alpha_i^t(m) \triangleq P\{Y_i^1 = y^1, ..., Y_i^t = y^t, S_i^t = m; \mathbf{x}_i^t\},$$

The initialization and recursion formulas are defined as

$$\begin{aligned}
\alpha_i^1(m) &= \delta_i(m) P[y^1 | S_i^1 = m, \mathbf{x}_i^t], \\
\alpha_i^{t+1}(m) &= \left[\sum_{\ell=1}^{M} \alpha_i^t(\ell) \mathbf{Q}_i(\ell, m; \mathbf{x}^t)\right] P[y^{t+1} | S_i^{t+1} = m, \mathbf{x}_i^t],
\end{aligned}$$

where $m=1,...,M$ and $t=1,...,T-1$. The likelihood for subject i can be expressed as $\sum_{m=1}^{M} \alpha_i^T(m)$ with the computation cost of $O(M^2T)$. Denote by $\boldsymbol{\theta} \triangleq [\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n]$, where $\boldsymbol{\theta}_i$ is the parameter vector for subject i. Therefore, the corresponding likelihood for all subjects becomes

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} \sum_{m=1}^{M} \alpha_i^T(m).$$

The maximum likelihood estimator $\hat{\boldsymbol{\theta}}_T = [\hat{\boldsymbol{\theta}}_{1,T}, ..., \hat{\boldsymbol{\theta}}_{n,T}]$ is constructed for each subject $i = 1, ..., n$.

The asymptotic consistency and normality of the maximum likelihood estimators for the hidden Markov modes have been established [Ler92; Bic98; DM01; Cap09; An13]. The following are common regularity assumptions in the literature:

(1) The dimension of the parameters is $\dim(\boldsymbol{\theta}^*) = c$, where c is known. Furthermore, we assume that $\boldsymbol{\theta}^*$ is an interior point of Θ which is a compact subset of $\mathbb{R}^c$.

(2) There exists $0 < \sigma_- \le \sigma_+ < \infty$ such that $\sigma_- \le P_{\boldsymbol{\theta}^*}\left(S^{t+1} = \ell \middle| S^t = m\right) \le \sigma_+$, for all $\ell, m \in \{1, \dots, M\}$ and $t = 1, \dots, T-1$; and $0 < \sum_{m=1}^{M} P_{\boldsymbol{\theta}^*}\left(Y^t = y \middle| S^t = m\right)$ for all $y \in \operatorname{supp} Y^t$ and $t = 1, \dots, T$.

(3) For each $\boldsymbol{\theta} \in \Theta$ the transition kernel indexed by $\boldsymbol{\theta}$ is Harris recurrent and aperiodic.

(4) The hidden Markov model is identifiable up to label-switching of the latent states.

(5) The transition kernel is continuous in $\boldsymbol{\theta}$ in an open neighborhood of $\boldsymbol{\theta}^*$.

(6) The log likelihood function is three times differentiable with respect to $\boldsymbol{\theta}$ and the derivatives are bounded.

(7) The Fisher information matrix $\mathbf{I}(\boldsymbol{\theta}^*) = E_{\boldsymbol{\theta}^*}\left\{\left[\frac{\partial}{\partial \theta} \log L_n(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial \theta} \log L_n(\boldsymbol{\theta})\right]^{\intercal}\right\}$ is positive definite.

Under assumptions (1) - (5),

$$\hat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta}^* \text{ almost surely as } T \to \infty.$$

If assumptions (6) - (7) also hold, we have,

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_i^*) \xrightarrow{d} \mathbf{N}[\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}^*)^{-1}] \text{ as } T \to \infty,$$

$$\text{where } \mathbf{I}(\boldsymbol{\theta}^*) = E_{\boldsymbol{\theta}^*}\left\{\left[\frac{\partial}{\partial \boldsymbol{\theta}} \log L_n(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial \boldsymbol{\theta}} \log L_n(\boldsymbol{\theta})\right]^{\intercal}\right\}.$$

The mixed hidden Markov model was proposed by [Alt07] to analyze the longitudinal data, which involves either numerical or Monte-Carlo integration of the random effects.

Those integration can be extremely slow, especially when the number of latent states and the sample size are large. The previous applications have considered relatively coarse time scales, e.g., daily or weekly measurements, and subsequently low data volume. In the context of wearable computing, subject activity is measured at least every minute and subsequently each has more than one-hundred thousand measurements in the observation period. Such volume allows for flexible modeling including estimation of distinct parameters for each subject; this is important as sensor data patterns can vary widely across subjects. Another advantage of the hidden Markov model approach is that it can not only detect the treatment effect on the mean activity counts but also on the transitions between lower and higher level activity states.

The validity of the hidden Markov model depends on the Markov assumption in the latent state process, i.e., $P(S_i^t|S_i^{t-1},...,S_i^1) = P(S_i^t|S_i^{t-1})$. Since the latent states are unobserved, the magnitude in the deviation from the Markov assumption is hard to check. The number of latent states can be selected using the minimum BIC criterion. The goodness-of-fit of the model can be checked by comparing the empirical versus the fitted distribution functions.

1.1.2 Hidden Semi-Markov Model

As is discussed the previous section, HMM is a bivariate discrete time stochastic process $\{S_i^t, Y_i{}^t\}_{t=1}^T$ for $i = 1,...,n$ such that

(1) $\{S_i^t\}$ is a Markov chain, i.e. $P(S_i^t|S_i^{t-1},...,S_i^1) = P(S_i^t|S_i^{t-1})$.

(2) $\{Y_i{}^t\}$ is a sequence of conditional independent random variables given $\{S_i^t\}$.

However, the Markovian assumption may be too strong to hold in practice. To overcome the lack of flexibility of HMM to model the temporal higher order dependence, the hidden semi-Markov model (HSMM) was proposed [Fer80; RM85; Lev86], which explicitly models the latent state durations rather than fixing them to be geometrically distributed as in HMM. So far, the application of hidden semi-Markov model in the context of wearable computing is for mobility tracking [Lee08; YK03] and activity recognition [NN07; Tho10].

The HSMM notations and model setup are exactly the same as HMM except that there is an additional component of latent state duration densities,

$$r_{m,i}(v;\mathbf{x}^t) \triangleq P\left(S_i^{t+1} = m, \ldots, S_i^{t+v} = m, S_i^{t+v+1} \neq m \middle| S_i^t = m, \mathbf{X}_i^t = \mathbf{x}^t\right),$$

for $v = 0, 1, \ldots$, $m = 1, \ldots, M$. A modified forward-backward recursion is adopted in order to compute the likelihood function for HSMM. Let $\boldsymbol{\theta}_i$ be the collection of subject-specific parameters for $i = 1, \ldots, n$. For each $t = 1, \ldots, T$, $m = 1, \ldots, M$ and $i = 1, \ldots, n$ define the forward-variables $\alpha_i^t(m;\boldsymbol{\theta}_i) = P_{\theta_i}\left(Y_i^1, \ldots, Y_i{}^t, S_i^t = m\right)$ where P_{θ_i} denotes probability under the distribution indexed by $\boldsymbol{\theta}_i$. We assume there is a finite latent state duration D; any state durations exceeding D are artificially censored. Then it can be shown [Rab89] that the forward-variables satisfy the following recursions

$$
\begin{aligned}
\alpha_i^1(m;\boldsymbol{\theta}_i) =& \delta_{m,i} r_{m,i}(1) P(Y_i^1|S_i^1 = m), \\
\alpha_i^2(m;\boldsymbol{\theta}_i) =& \delta_{m,i} r_{m,i}(2) \prod_{\tau=1}^{2} P(Y_i^\tau|S_i^\tau = m) + \\
& \sum_{\ell \neq m} \alpha_i^1(\ell;\boldsymbol{\theta}_i) Q_i(\ell,m) r_{m,i}(1) P(Y_i^2|S_i^2 = m), \\
\alpha_i^3(m;\boldsymbol{\theta}_i) =& \delta_{m,i} r_{m,i}(3) \prod_{\tau=1}^{3} P(Y_i^\tau|S_i^\tau = m) + \\
& \sum_{d=1}^{2} \sum_{\ell \neq m} \alpha_i^{3-d}(\ell;\boldsymbol{\theta}_i) Q_i(\ell,m) r_{m,i}(d) \prod_{\tau=4-d}^{3} P(Y_i^\tau|S_i^\tau = m), \\
& \vdots \\
\alpha_i^D(m;\boldsymbol{\theta}_i) =& \delta_{m,i} r_{m,i}(D) \prod_{\tau=1}^{D} P(Y_i^\tau|S_i^\tau = m) + \\
& \sum_{d=1}^{D-1} \sum_{\ell \neq m} \alpha_i^{D-d}(\ell;\boldsymbol{\theta}_i) Q_i(\ell,m) r_{m,i}(d) \prod_{\tau=D+1-d}^{D} P(Y_i^\tau|S_i^\tau = m),
\end{aligned}
$$

where the terms on the right-hand-side of the above equations are dictated by $\boldsymbol{\theta}_i$ through the models posited previously. For $t = D+1,\ldots,T$ the forward-variables satisfy

$$
\alpha_i^t(m;\boldsymbol{\theta}_i) = \sum_{\ell=1}^{M} \sum_{d=1}^{D} \left[\alpha_i^{t-d}(\ell;\boldsymbol{\theta}_i) Q_i(\ell,m) r_{m,i}(d) \prod_{\tau=t-d+1}^{t} P(Y_i^\tau|S_i^\tau = m) \right].
$$

Thus, the likelihood for subject i is $\mathscr{L}_i(\boldsymbol{\theta}_i) = \sum_{m=1}^{M} \alpha_i^T(m;\boldsymbol{\theta}_i)$ with the computation cost of $O(M^2 D^2 T)$. Therefore, the corresponding likelihood for all subjects becomes

$$
L_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} \sum_{m=1}^{M} \alpha_i^T(m;\boldsymbol{\theta}_i).
$$

The maximum likelihood estimator $\hat{\boldsymbol{\theta}}_T = [\hat{\boldsymbol{\theta}}_{1,T},...,\hat{\boldsymbol{\theta}}_{n,T}]$ is constructed for each subject $i = 1,...,n$ using either the Baum-Welch algorithm [Bau70] or some gradient-based approach like the Quasi-Newton method.

The asymptotic consistency and normality of the maximum likelihood estimators for the HSMM have been shown in [BL09] and [TL11]. In addition to the regularity conditions (D1) - (D7), the state duration distribution $r_{m,i}(\cdot)$ should have finite support for all $m \in \{1,\ldots,M\}$

(8). Then, under (1) - (5) and (8),

$$\hat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta}^* \text{ almost surely as } T \to \infty.$$

If assumptions (6) - (7) also hold, we have,

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^*) \xrightarrow{d} \mathbf{N}[\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}^*)^{-1}] \text{ as } T \to \infty,$$

$$\text{where } \mathbf{I}(\boldsymbol{\theta}^*) = E_{\boldsymbol{\theta}^*}\left\{\left[\frac{\partial}{\partial \boldsymbol{\theta}} \log L_n(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial \boldsymbol{\theta}} \log L_n(\boldsymbol{\theta})\right]^{\intercal}\right\}.$$

In the HSMM framework, no data aggregation or transformation is needed so that there is no loss of information. The mixture of a zero-inflated Poisson and Poisson distributions is suitable to model the highly right-skewed activity count data. Moreover, the HSMM is capable of capturing treatment effects not only on the mean activity counts but also on the latent state durations and transitions. In this thesis, we will develop the novel framework of a subject-specific zero-inflated Poisson hidden semi-Markov model (in Chapter 2) and a continuous-time hierarchical hidden Markov model (in Chapter 3) to characterize activity profiles from accelerometer data, along with methods for efficient parameter estimation and theories for statistical inference.

1.1.3 Markov decision processes

In the foregoing discussion of various types of Markov models, the analysis goal is to estimate the parameters indexing each component in the model, including the transition probabilities and the conditional distributions of observations. However, if the goal is to actively engage in controlling the system dynamics of a Markov model so as to maximize some prespecified outcome, we need to utilize a Markov decision process framework to be introduced in this subsection.

We begin with the simplest form of a fully observable Markov decision process, which can be described as a tuple $(\mathscr{S}, \mathscr{A}, Q, U)$, where: $\mathscr{S} \triangleq \{1, \ldots, M\}$ is a finite set of states; $\mathscr{A} \triangleq \{1, \ldots, L\}$ is a finite set of actions; $Q : \mathscr{S} \times \mathscr{A} \to \mathscr{P}_S(\mathscr{S})$ is a state transition function for each state and action, where $\mathscr{P}_s(\cdot)$ is a probability distribution over states; and $U : \mathscr{S} \times \mathscr{A} \to \mathbb{R}$ is a utility function, which is also known as the expected immediate reward obtained by taking each action in each state. A policy is a sequence of decision rules $\boldsymbol{\pi} \triangleq \{\pi^j\}_{j \geq 1}$ with $\pi^j : \mathscr{S} \to \mathscr{P}_A(\mathscr{A})$, where $\mathscr{P}_A(\cdot)$ is a probability distribution over actions. Let $\mathbb{E}^{\pi}$ denote the probability distribution induced by following the actions given by $\boldsymbol{\pi}$. We consider the following two measures of cumulative utility that evaluates the goodness of a given policy:

(i) discounted mean utility

$$V_{\text{dis}}(\boldsymbol{\pi}) \triangleq \mathbb{E}^{\pi} \sum_{j\geq 1} \left(\gamma^{j-1} U^j\right),$$

where $\gamma \in [0, 1)$ is a discount factor; and (ii) average utility

$$V_{\text{ave}}(\boldsymbol{\pi}) \triangleq \lim_{J\to\infty} \mathbb{E}^{\pi} \left(\frac{1}{J} \sum_{j=1}^{J} U^j \right).$$

These two cumulative measures are used almost exclusively in indefinite decision problems [Pow07; Bus17; SB18]. Denote by $V(\boldsymbol{\pi})$ without a subscript to generically denote either of these cumulative utility measures. Then $\boldsymbol{\pi}^{\text{opt}}$ is optimal with respect to V if $V(\boldsymbol{\pi}^{\text{opt}}) \geq V(\boldsymbol{\pi})$ for all $\boldsymbol{\pi}$.

Define the discounted state-value function $v_{\text{dis}}(\pi, \mathbf{s}) = \mathbb{E}^{\pi} \left(\sum_{j\geq 0} \gamma^j U^{t+j} \middle| \mathbf{S}^t = \mathbf{s} \right)$ which does not depend on t because the Markov decision process is assumed to be homogeneous. For any distribution R on $\mathscr{S}$, termed a reference distribution, define $V_{\text{dis}}^{R}(\pi) = \int v_{\text{dis}}(\pi, \mathbf{s}) dR(\mathbf{s})$ then $V_{\text{dis}}(\pi) = V_{\text{dis}}^{R_0}(\pi)$, where R_0 is the initial state distribution. Because R_0 is unknown, one might take the empirical distribution of $\mathbf{S}^1$ or some other reference distribution constructed from historical data [see Luc19]. Define $\pi^{\text{opt},R} = \arg\max_{\pi\in\Pi} V_{\text{dis}}^{R}(\pi)$, where Π denotes a class of regimes of interest. In the discounted reward case it can be shown [e.g., Luc19] that the state-value function satisfies the following recursion

$$0 = \mathbb{E}\left[\frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j \middle| \mathbf{S}^j)} \left\{ U^j + \gamma\, v_{\text{dis}}(\boldsymbol{\pi}, \mathbf{S}^{j+1}) - v_{\text{dis}}(\boldsymbol{\pi}, \mathbf{S}^j) \right\} \phi(\mathbf{S}^j) \right], \tag{1.1}$$

for all j and any $\phi : \mathscr{S} \to \mathbb{R}^d$, where the ratio $\pi_{A^j}(\mathbf{S}^j)/P(A^j|\mathbf{S}^j)$ is an importance sampling weight.

An estimating equation for the average reward setting is derived using a similar strategy to the discounted case. For each π define the differential value

$$\delta(\boldsymbol{\pi}, \mathbf{s}) \triangleq \lim_{J\to\infty} \mathbb{E}^{\pi} \left[\sum_{j=1}^{J} \left\{ U^j - V_{\text{ave}}(\boldsymbol{\pi}) \right\} \middle| \mathbf{S}^1 = \mathbf{s} \right],$$

which is well-defined under regularity conditions provided in Chapter 4. Then it can be shown [e.g., Put14; Mur16; Lia19] that the $V_{\text{ave}}(\boldsymbol{\pi})$ satisfies the recursion

$$0 = \mathbb{E}\left[\frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j|\mathbf{S}^j)} \left\{ U^j - V_{\text{ave}}(\boldsymbol{\pi}) + \delta(\boldsymbol{\pi}, \mathbf{S}^{j+1}) - \delta(\boldsymbol{\pi}, \mathbf{S}^j) \right\} \psi(\mathbf{S}^j) \right], \tag{1.2}$$

for all j and any $\psi : \mathscr{S} \to \mathbb{R}^e$.

1.1.4 Partially observable Markov decision process

When the states are not fully observable, POMDP provides an elegant and systematic method to address this problem. POMDPs have long been studied in the computer science literature with applications in robotics, scheduling, videogames, and wildlife management [see, for example, Kae98; Cas98; JP03; Han98b; Ji07; SB18]. The first papers on POMDP came out in 1960s [How60; Dra62; Ast65]. Earlier research on POMDP focused on exact solution algorithms, including one-pass algorithm [Son71], policy iteration [Son78], enumeration algorithm [Mon82], linear support algorithm [Che88], witness algorithm [Lit94], and incremental pruning [ZL96; Kae98; Cas13]. However, as exact solutions quickly become intractable in POMDP problems with large state spaces, there has been a growing interest in approximate solution algorithms, including finite state controller policy iteration [Han98a], grid based value iteration [ZH01], point based value iteration [JP03], point based policy iteration [Ji07], heuristic search value iteration [SS12], and Monte-Carlo tree search [SV10; Bai14].

POMDP can be described as a tuple $\{\mathscr{S}, \mathscr{A}, Q, U, \mathscr{X}, G\}$. The definition of $\mathscr{S}$, $\mathscr{A}$, Q, and U are the same as in a Markov decision process. However, the states are latent and $\mathscr{X}$ is the set of observations available. Moreover, $G : \mathscr{S} \times \mathscr{A} \to \mathscr{P}_X(\mathscr{X})$ is an observation function for each action and resulting state, where $\mathscr{P}_X(\cdot)$ is a probability distribution over observations. In the example of estimating an optimal rule for treating bipolar disorder patients, $\mathscr{S}$ is the set of latent mood states; $\mathscr{A}$ is the set of treatment options; Q determines the transition of the latent mood states; U measures the goodness of selecting each action in each state; $\mathscr{X}$ is the set of answers provided by the patients in the mood-related questionnaire; G specifies the conditional distribution of answers given mood state and treatment.

Assume the latent state evolves according to a Markov process indexed by (1) an initial state distribution, $\delta_m \triangleq P\{S(T^1) = m\}$ for $m = 1, \ldots, M$, where $T^1 \equiv 0$; (2) treatment-dependent transition probabilities $P_{m,\ell}(t; a) \triangleq P\{S(T^j) = \ell | T^{j+1} - T^j = t, S(T^j) = m, A^j = a\}$ for $m, \ell = 1, \ldots, M$, and $j = 1, \ldots, J-1$. The continuous-time transition probabilities can be parameterized by a transition rate matrix, $Q(a) \triangleq \{q_{m,\ell}(a)\}_{m,\ell=1,\ldots,M} \in \mathbb{R}^{M \times M}$, which is also known as the 'infinitesimal generator' [Pyk61a; Pyk61b; Alb62]. The diagonals in the generator matrix are $q_{m,m}(a) \triangleq -\lim_{t\to 0^+}[P\{S(T^{j+1}) \neq m | T^{j+1} - T^j = t, S(T^j) = m, A = a\}/t]$ and the off-diagonals are $q_{m,\ell}(a) \triangleq \lim_{t\to 0^+}[P\{S(T^{j+1}) = \ell | T^{j+1} - T^j = t, S(T^j) = m, A = a\}/t]$, which implies $q_{m,m}(a) = -\sum_{\ell \neq m} q_{m,\ell}$ for $m = 1, \ldots, M$. The transition probabilities can be expressed in terms of the generator matrix by $P_{m,\ell}(t; a) = [\exp\{t \cdot Q(a)\}]_{m,\ell}$, which is the

$(m,\ell)^{th}$ entry in the matrix exponential.

Define the overline notation $\overline{A}^j \triangleq \{A^1,\dots,A^j\}$ and $\overline{\mathbf{X}}^j \triangleq \{\mathbf{X}^1,\dots,\mathbf{X}^j\}$. Define $\overline{\Delta}^j \triangleq \{\Delta^2,\dots,\Delta^j\}$ for $j \geq 2$, where $\Delta^j \triangleq T^j - T^{j-1}$. Let $\mathbf{B}^k \in \mathbb{R}^M$ be a probability distribution over the latent mood states conditioning on $\{\overline{A}^{j-1},\overline{\mathbf{X}}^j,\overline{\Delta}^j\}$, where $\overline{\Delta}^1 \triangleq \emptyset, \overline{A}^0 \triangleq \emptyset$. The $\mathbf{B}^k$'s are known as belief states in the POMDP literature [Son78; Kae98]. This means $\mathbf{B}^j = [B_1^j,\dots,B_m^j]$, where $B_m^j \triangleq P\{S(T^j) = m|\overline{A}^{j-1},\overline{\mathbf{X}}^j,\overline{\Delta}^j\}$. In the proposed latent Markov process, $\{\mathbf{B}^j,\mathbf{X}^j\}$ is a sufficient statistic for $\{\overline{A}^{j-1},\overline{\mathbf{X}}^j,\overline{\Delta}^j\}$. This shows that given the current belief state and observation, no additional data about the past treatment, interarrival times, and observations would provide further information about the history. Therefore, it makes sense to construct the policies based on $\{\mathbf{B}^j,\mathbf{X}^j\}$ in the POMDP framework, because the entire history of disease evolution trajectory can be parsimoniously summarized by the current belief state and observation. By replacing $\mathbf{S}^j$ with $\{\mathbf{B}^j,\mathbf{X}^j\}$ in equations (1.1) and (1.2), we can obtain the estimating equations that define the value for any arbitrary fixed policy in the POMDP for both discounted utility and average utility setting. Essentially, the POMDP problem is reduced to a continuous state Markov decision process problem, where the redefined partially observable states can be estimated using methodologies in the HMM literature. This is exactly the estimating equation based approach we develop in Chapter 4 to identify the optimal treatment policy in a bipolar disorder study.

1.2 Contributions and outline

In Chapter 2, the motivating data come from a randomized cross-over study designed to evaluate the effectiveness of meloxicam, a nonsteroidal anti-inflammatory drug (NSAID), in owned, pet adult domestic cats with OA-associated pain [Gru15] over an 11-week period. In this study, cat activity patterns were measured at one minute intervals using an omnidirectional accelerometer [see Las08, for accelerometer specifications]. Accelerometer readings are integers quantifying the intensity of movement over the preceding, pre-defined epoch. Thus, for each subject, the accelerometer produces high-frequency, integer-valued, longitudinal data. One common approach for the analyses of such data is to aggregate the observed accelerometer within each subject and treatment condition and subsequently to use an ANOVA to compare treatment with control [Las08; Gru15; Gru16]. Alternatively, one could aggregate the data over a shorter time interval, e.g., a day, and model the aggregated process using methods for smooth longitudinal data [Gru17]. However such aggregation can obscure changes in behavior patterns which do not produce a large difference in mean activity. For example, an effective treatment that reduces pain may lead to higher quality rest (more zero activity readings) *and* more high-intensity movement (more high activity readings) producing only a small change in mean activity. High volume activity data has the potential to be a more sensitive outcome measure, but thus far, methods for analysis of such complex, high-volume multidirectional effects on activity, that takes into account the wide variation in individual activity patterns, has not been developed in humans or animal models.

We propose to model the minute-by-minute accelerometer data using a hidden semi-Markov model. Such a model is scientifically appealing in that the hidden states can be viewed as corresponding to latent (unobserved) activities, e.g., running, walking, napping, etc., and state duration corresponds to the length of time a subject is engaged in a given activity. Latent state-space models are common in the analysis of mobility data measured using wearable computing [see He07; MS10; Nic11; LC11; RC14, for related models in the context of human mobility data]. In the context of treatment evaluation, hidden Markov models have also been used to model latent health states and subsequently conduct inference for activity patterns in terms of transitions among these states and activity patterns within each state [Sco05; Shi10]; however, previous applications have considered relatively coarse time scales, e.g., daily or weekly measurements, and subsequently low data volume. In the current application, subject activity is measured every minute and as a result each subject has more than one-hundred thousand measurements during the observation period. Such high volume allows for flexible modeling and estimation of distinct parameters for each subject; this is important as activity patterns can vary widely across subjects. In the current example of the spontaneous cat OA model, more than 70% of the observations are zero and therefore we propose a zero-inflated hidden Poisson semi-Markov model with patient-specific intensities in each latent state. The proposed model is an extension of a zero-inflated Poisson hidden Markov model [DB11; OR12] to the hidden semi-Markov model framework that allows for distinct state duration densities [Fer80; RM85; Lev86]. Furthermore, to facilitate computationally efficient estimation with high-frequency data, we propose a two stage estimation procedure [Han74; SD97] that can be used with millions of observations without specialized computing resources.

In Chapter 3, our work is motivated by the physical activity data set from the 2003 - 2004 National Health and Nutrition Examination Survey (NHANES). In this study, human activity patterns were measured at one-minute intervals for up to seven days using the ActiGraph Model 7164 accelerometer [Tro08; Met08; Sch15]. Activity for each minute was recorded as an integer-valued intensity-level commonly referred to as an activity count. In the study, subjects were instructed to remove the device during sleep or while washing (to keep it dry). Therefore, the observed data comprise high-frequency, integer-valued activity counts for each subject with intervals of missing values corresponding to when the device was removed. Prior work on modeling activity counts has focused on aggregation and other smoothing techniques. One common approach is to average the activity counts over time for each subject and then compare the group means using two-sample t-tests [Tro08], analysis of covariance [Han12], or linear mixed effects models [Cra04]. However, in these approaches, averaging focuses on overall activity levels and may obscure trends in activity type, activity duration, and transitions between activities. Another approach is to use functional data analysis methods wherein the integer activity counts are first log-transformed to fix the right-skewedness in the distribution and the transformed activity modeled as a function of time of day and other covariates [Mor06; Xia14; Gru17]. These approaches are best-suited for the identification of smooth, cyclical patterns in the data whereas the observed accelerometer data are characterized by abrupt (i.e., non-smooth) changes in activity levels. Discrete-time hidden Markov models are another common approach to the analysis of mobility data measured by wearable devices [He07; Nic11; RC14; Wit14]. A technical limitation of the discrete-time approach, is that it assumes that the observations are equally spaced in time.

We propose to model minute-by-minute accelerometer data using a hierarchical continuous-time hidden Markov model (HCTHMM). This model is aligned with scientific models of activity as the latent states represent different types of unobserved physical activities. The continuous-time Markov process for the latent states evolution avoids having to perform imputation for missing yet allows for the possibility that the temporal measurements are irregularly spaced. Furthermore, the proposed model can incorporate both baseline subject covariates and time-varying environmental factors. The proposed model is hierarchical in that it is parameterized by: (i) subject-specific parameters to account for variability between subjects; (ii) subgroup-specific parameters parameters to account for similarity in activity patterns within groups; (iii) population parameters that are common across all subjects. This specification allows us to pool information on some parameters while retaining between-group and between-subject variability. We proposed an estimator of these parameters that is based on consensus optimization using the alternating direction method of multipliers (ADMM). There is a vast literature on the convergence properties of ADMM [Boy11; Shi14; HL17] which can be readily ported to the proposed algorithm. Finally, we use the nonparametric bootstrap [Efr92] to estimate the sampling distributions of parameter estimators and to conduct statistical inference.

In Chapter 4, we focus on the methodology for estimating optimal treatment policy in observational longitudinal studies, which are a common means to study treatment efficacy and safety in chronic mental illness. In many such studies, treatment changes may be initiated either by the patient or their clinician and can thus vary widely across patients in their timing, number, and type. Indeed, in the observational longitudinal pathway of the STEP-BD study of bipolar depression, one of the motivations for this work, no two patients have the same treatment history even after coarsening clinic visits to a weekly time-scale. Estimation of an optimal treatment regime using such data is challenging as one cannot naively pool together patients with the same treatment history, as required by methods based on inverse probability weighting [Rob99; Mur01; Laa06; Rob08], nor is it possible to apply backwards induction over the decision points, as is done Q-learning and its variants [Mur05; Moo07; Hen10; Sch14; Moo14a; Son15; Tay15; Zha18; Ert19].

Thus, additional structure is needed to effectively pool information across patients and within a patient over time. Current scientific theory for many chronic mental illnesses maintains that a patient's disease status can be conceptualized as transitioning among a small number discrete states. We use this theory to inform the construction of a POMDP framework of patient health trajectories wherein observed health outcomes are dictated by a patient's latent health state. Using this model, we derive an estimator of an optimal treatment regime under two common paradigms for quantifying long-term patient health. The finite sample performance of the proposed estimator is demonstrated through a series of simulation experiments and application to the observational pathway of the STEP-BD study. We find that the proposed method provides high-quality estimates of an optimal treatment strategy in settings where existing approaches cannot be applied without *ad hoc* modifications.

CHAPTER

2

HIDDEN MARKOV MODEL FOR CAT ACCELEROMETER DATA

2.1 Introduction

To anyone who suffers chronic and persistent musculoskeletal pain, the negative impact on their quality of life is constant. In the U.S., more than 100 million people (nearly one third of the population) suffer from persistent pain with an associated economic cost of $600 billion USD annually; this cost exceeds that of cardiovascular disease, cancer, and diabetes combined [Cou11]. OA results in the deterioration of all components of joints [Buc04] and is often associated with pain. OA-associated pain results in significant morbidity and economic costs [Che12]; these costs, an aging population, and the growing knowledge of the health and psychological benefits of maintaining mobility and activity, have made the treatment of OA and the associated pain a public health priority [Had85; Eld00; Hoo12; McC16]. However, recent review papers show that the current practice of translational biomedical research is not producing new therapeutics for pain control in humans [Vie08; Mog09; Hay14]. While these reviews highlight the lack of translation of basic research into new approved therapeutics for treatment of persistent pain in humans, they also discuss how the processes involved could be optimized to improve the chances of successful translation, including discussion of improved models and more relevant outcome measures. As OA and the associated pain affects all mammals, the study of OA in non-human animals is both important in its own right (to increase the function and quality of life in affected animals) and for its ability to generate new knowledge about the treatment of OA in humans. Indeed, recently, the use of naturally occurring OA in pet animals has been suggested as a means of helping to improve the translational research for the development of analgesic therapeutics in humans [Las17]. A significant advantage of naturally occurring disease in pets as a model of human conditions is the variation and complexity of the model. Measurement of spontaneous activity as an outcome measure may be particularly relevant to translational work as spontaneous activity may relate to spontaneous pain - something that has been difficult to model in animals [Par18]. Just as the variability in naturally occurring disease is an advantage, the variability in clinically meaningful outcome measures such as activity is a challenge. This could be the reason for the lack of use of activity as a functional outcome measure in pain studies in humans, despite the crucial importance of mobility and activity to quality of life [McC16]. The work we are reporting is motivated by our involvement in a study of OA-associated pain and the treatment effects on mobility and activity in a naturally occurring OA model in domestic cats [Las10; Sli11]. The methodology we propose presents a significant advance in how such activity data are analyzed when used as an outcome measure and has relevance to both large animal models and clinical evaluation in humans.

The impetus behind the current work is a randomized cross-over study designed to evaluate the effectiveness of meloxicam, a nonsteroidal anti-inflammatory drug (NSAID), in owned, pet adult domestic cats with OA-associated pain [Gru15] over an 11-week period. In this study, cat activity patterns were measured at one minute intervals using an omnidirectional accelerometer [see Las08, for accelerometer specifications]. Accelerometer readings are integers quantifying the intensity of movement over the preceding, pre-defined epoch. Thus, for each subject, the accelerometer produces high-frequency, integer-valued, longitudinal data. One common approach for the analyses of such data is to aggregate the observed accelerometer within each subject and treatment condition and subsequently to use an ANOVA to compare treatment with control [Las08; Gru15; Gru16]. Alternatively, one could aggregate the data over a shorter time interval, e.g., a day, and model the aggregated process using methods for smooth longitudinal data [Gru17]. However such aggregation can obscure changes in behavior patterns which do not produce a large difference in mean activity. For example, an effective treatment that reduces pain may lead to higher quality rest (more zero activity readings) *and* more high-intensity movement (more high activity readings) producing only a small change in mean activity. High volume activity data has the potential to be a more sensitive outcome measure, but thus far, methods for analysis of such complex, high-volume multidirectional effects on activity, that takes into account the wide variation in individual activity patterns, has not been developed in humans or animal models.

We propose to model the minute-by-minute accelerometer data using a hidden semi-Markov model. Such a model is scientifically appealing in that the hidden states can be viewed as corresponding to latent (unobserved) activities, e.g., running, walking, napping, etc., and state duration corresponds to the length of time a subject is engaged in a given activity. Latent state-space models are common in the analysis of mobility data measured using wearable computing [see He07; MS10; Nic11; LC11; RC14, for related models in the context of human mobility data]. In the context of treatment evaluation, hidden Markov models have also been used to model latent health states and subsequently conduct inference for activity patterns in terms of transitions among these states and activity patterns within each state [Sco05; Shi10]; however, previous applications have considered relatively coarse time scales, e.g., daily or weekly measurements, and subsequently low data volume. In the current application, subject activity is measured every minute and as a result each subject has more than one-hundred thousand measurements during the observation period. Such high volume allows for flexible modeling and estimation of distinct parameters for each subject; this is important as activity patterns can vary widely across subjects. In the current example of the spontaneous cat OA model, more than 70% of the observations are zero and therefore we propose a zero-inflated hidden Poisson semi-Markov model with patient-specific intensities in each latent state. The proposed model is an extension of a zero-inflated Poisson hidden Markov model [DB11; OR12] to the hidden semi-Markov model framework that allows for distinct state duration densities [Fer80; RM85; Lev86]. Furthermore, to facilitate computationally efficient estimation with high-frequency data, we propose a two stage estimation procedure [Han74; SD97] that can be used with millions of observations without specialized computing resources.

2.2 Study of the effect of NSAID-related pain relief in pet cats with OA-associated pain

The data motivating this work come from a study of the treatment effects of meloxicam (a non-steroidal anti-inflammatory drug) in a feline model of spontaneous OA-associated pain [Gru14; Gru15]. This study follows a randomized, double-blind, placebo-controlled, crossover study to evaluate the effectiveness of meloxicam treatment to improve mobility in owned, pet cats with OA-associated pain. This study was approved by the Animal Care and Use Committee (Protocol # 11-102-O) at North Carolina State University College of Veterinary Medicine (NCSU-CVM), and written owner consent was granted in each case following verbal discussion of the study. A detailed description of the study protocol can be found in [Gru15].

Figure 2.1 displays the daily activity profile for a typical subject. There are multiple prolonged intervals in which there is no movement. In total, more than 70% of the minute-by-minute activity counts are zero; furthermore, the non-zero counts are heavily right-skewed owing to infrequent but high-intensity activities. Thus, such cat accelerometer data are heavy-tailed with an excessive proportion of zeros.

A total of 66 subjects were enrolled in the study among which 58 had available accelerometer measurements from both arms of the cross-over design. Subjects were given placebo during the open label baseline period (Weeks 1 and 2). In the first blinded treatment period (Weeks 3, 4, and 5) half of the subjects were randomized to receive meloxicam and the remainder were randomized to receive placebo. At the end of the first blinded phase, subjects entered a three week blinded washout (Weeks 6,7, and 8) before switching treatments for the second blinded phase. All subjects wore an accelerometer (Actical) on their collar throughout the entire study, with the epoch set at 1 minute, providing an activity count for every minute of the 11 week study for each subject.

Figure 2.2 displays mean activity counts for each subject across each period of the study. In this figure, each gray line shows the mean activity count for a single subject in the study, and the black line shows the overall mean activity count for all subjects in their associated treatment group. The left panel shows data for the subjects that were randomized to meloxicam in the first treatment period, whereas the right panel shows data for the subjects that were randomized to placebo in the first treatment period. In both panels, one can see that the overall mean activity count is higher under meloxicam than placebo. However, there is large between-subject variability which we will model using subject-level effects as described in the following section.

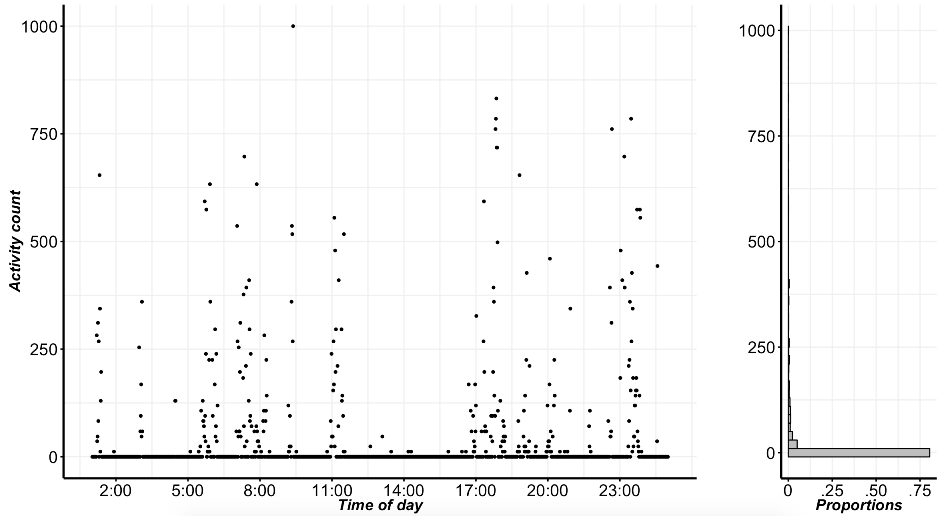

Figure 2.1 Figure 1: **Left:** Sample daily accelerometer data for a typical cat in the study of the treatment effect of meloxicam. There are multiple prolonged intervals in which the cat does not move at all.

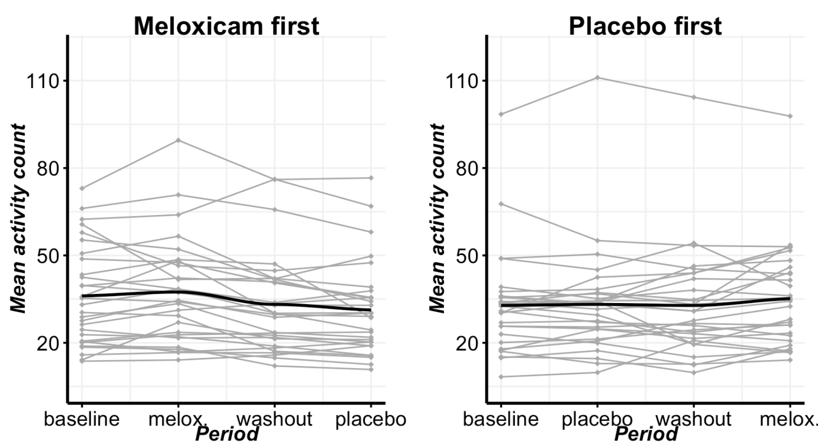

Figure 2.2 Mean activity count over each period for each subject separated by treatment group in the study.

2.3 Zero-inflated Poisson hidden semi-Markov model

We assume that the observed data are of the form $\left\{\left(\mathbf{W}_i, \mathbf{Z}_i^1, \mathbf{X}_i^1, Y_i^1, \ldots, \mathbf{Z}_i^T, \mathbf{X}_i^T, Y_i^T\right)\right\}_{i=1}^n$ and comprise n independent replicates of the trajectory $\left(\mathbf{W}, \mathbf{Z}^1, \mathbf{X}^1, Y^1, \ldots, \mathbf{Z}^T, \mathbf{X}^T, Y^T\right)$, where: $\mathbf{W} \in \mathbb{R}^d$ denotes baseline (pre-treatment) subject characteristics; $T \in \mathbb{N}_+ = \{1, 2, \ldots\}$ is a fixed time horizon; $\mathbf{Z}^t \in \mathbb{R}^p$ denotes environmental factors at $t = 1, \dots, T$; $\mathbf{X}^t \in \mathbb{R}^q$ encodes treatment administered at time $t = 1, \dots, T$; and $Y^t \in \mathbb{N} = \{0\} \bigcup \mathbb{N}_+$ denotes the accelerometer activity reading at time $t = 1, \dots, T$. In our application, $\mathbf{Z}^t \in \{0, 1\}^p$ comprises indicators coding period of day and weekend versus weekday, and $\mathbf{X}^t \in \{0, 1\}$ encodes treatment indicators (active versus placebo). While we develop our models allowing for rather general environmental factors and treatment processes, having binary factors in our application makes it simpler to enforce a stochastic ordering on the distributions indexed by latent states which is important for interpretability and coherence of the fitted models, e.g., if one wants the latent states ordered by mean activity intensity within each subject *and* to have these latent states align across subjects. We model the evolution of the accelerometer data using a zero-inflated Poisson hidden semi-Markov model which we develop in the remainder of this section.

Let $S_i^t \in \mathscr{S} = \{1, \dots, M\}$ denote the latent state, i.e., the unobserved activity type, of subject $i = 1, \dots, n$ at time $t = 1, \dots, T$. This latent state evolves according to a semi-Markov model indexed by: (i) an initial state distribution, $\delta_{m,i} \triangleq P(S_i^1 = m)$ for $m = 1, \dots, M$; (ii) state duration distributions $r_{m,i}(v; \mathbf{x}^t, \mathbf{z}^t) \triangleq P\big(S_i^{t+1} = m, \dots, S_i^{t+v} = m, S_i^{t+v+1} \neq m \big| S_i^t = m, \mathbf{X}_i^t = \mathbf{x}^t, \mathbf{Z}_i^t = \mathbf{z}^t\big)$ for $v \geq 0$, $m = 1, \dots, M$; and (iii) the transition probabilities $Q_i(m, \ell; \mathbf{x}^t, \mathbf{z}^t) = P\big(S_i^{t+1} = \ell \big| S_i^t = m, S_i^{t+1} \neq m, \mathbf{X}_i^t = \mathbf{x}^t, \mathbf{Z}_i^t = \mathbf{z}^t\big)$ for $m, \ell = 1, \dots, M$ [see Yu15, and references therein]. We model the initial state distribution nonparametrically using sample frequencies of the estimated latent states (see below). We assume a multinomial logistic regression model for the transition probabilities so that

$$Q_i(m, \ell; \mathbf{x}^t, \mathbf{z}^t) = \frac{\exp(\mathbf{d}'_{m,\ell,i}\mathbf{x}^t + \varrho'_{m,\ell,i}\mathbf{z}^t)}{\sum_{k=1}^M \exp(\mathbf{d}'_{m,k,i}\mathbf{x}^t + \varrho'_{m,k,i}\mathbf{z}^t)},$$

where $\mathbf{d}_{m,k,i}$ and $\varrho_{m,k,i}$ are unknown coefficient vectors for $m, k = 1, \dots, M$, $i = 1, \dots, n$. We assume that the state duration distribution follows a latent accelerated failure time model so that

$$r_{m,i}(v; \mathbf{x}^t, \mathbf{z}^t) = \exp(\mathbf{c}'_{m,i}\mathbf{x}^t + \eta'_{m,i}\mathbf{z}^t) \int_v^{v+1} f_{m,i}\{\exp(\mathbf{c}'_{m,i}\mathbf{x}^t + \eta'_{m,i}\mathbf{z}^t)u\} d\,u,$$

where $f_{m,i}$ is a base-density corresponding to a null treatment condition (coded as $\mathbf{x}^t \equiv 0$) and $\mathbf{c}_{m,i}$, $\eta_{m,i}$ are unknown coefficients vectors for $m = 1, \dots, M$, $i = 1, \dots, n$. We further assume that $Y_i^t \perp (\mathbf{Z}_i^1, \mathbf{X}_i^1, S_i^1, \dots, \mathbf{Z}_i^{t-1}, \mathbf{X}_i^{t-1}, S_i^{t-1}) \big| (\mathbf{Z}_i^t, \mathbf{X}_i^t, S_i^t)$ and that the accelerometer activity readings are distributed as

$$
\begin{aligned}
&P(Y_i^t = y \,\big|\, S_i^t = 1, \mathbf{Z}_i^t = \mathbf{z}^t, \mathbf{X}_i^t = \mathbf{x}^t) \\
=&p_i^t(\mathbf{z}^t, \mathbf{x}^t) 1_{y^t=0} + \left\{1 - p_i^t(\mathbf{z}^t, \mathbf{x}^t)\right\} \frac{\{\lambda_i^t(m, \mathbf{z}^t, \mathbf{x}^t)\}^y e^{\{\lambda_i^t(m, \mathbf{z}^t, \mathbf{x}^t)\}}}{y!},
\end{aligned}
$$

when the subject is in the first latent state, and

$$
P\left(Y_i^t = y \,\big|\, S_i^t = m, \mathbf{Z}_i^t = \mathbf{z}^t, \mathbf{X}_i^t = \mathbf{x}^t\right) = \frac{\{\lambda_i^t(m, \mathbf{z}^t, \mathbf{x}^t)\}^y \exp\{\lambda_i^t(m, \mathbf{z}^t, \mathbf{x}^t)\}}{y!},
$$

when the subject is in latent states $m = 2, 3, \ldots, M$. Thus, accelerometer readings are modeled a zero-inflated Poisson with intensity functions $\lambda_i^t : \mathscr{S} \times \mathbb{R}^q \times \mathbb{R}^p \to \mathbb{R}_+$ and weight functions $p_i^t : \mathscr{S} \times \mathbb{R}^q \times \mathbb{R}^p \to [0,1]$ for $i = 1, \ldots, n$, $t = 1, \ldots, T$. Furthermore, for each subject i, time t, and latent state m, we assume that these functions are of the form

$$
\begin{aligned}
\log\left\{\frac{p_i^t(\mathbf{z}^t, \mathbf{x}^t)}{1 - p_i^t(\mathbf{z}^t, \mathbf{x}^t)}\right\} &= b_{0,0,i} + \mathbf{b}'_{1,0,i}\mathbf{x}^t + \gamma'_{0,i}\mathbf{z}^t, \\
\log\left\{\lambda_i^t(m, \mathbf{z}^t, \mathbf{x}^t)\right\} &= b_{0,m,i} + \mathbf{b}'_{1,m,i}\mathbf{x}^t + \gamma'_{m,i}\mathbf{z}^t,
\end{aligned}
$$

where $b_{0,k,i}, \mathbf{b}_{1,k,i}, \gamma_{k,i}$, $k = 0, 1, \ldots, M$ are unknown coefficient vectors. In our application, in which $\mathbf{x}^t$ and $\mathbf{z}^t$ are represented as binary vectors, we impose the constraints $b_{0,m+1,i} > b_{0,m,i}$, $\mathbf{b}_{1,m+1,i} \geq \mathbf{b}_{1,m,i}$, and $\boldsymbol{\gamma}_{m+1,i} \geq \boldsymbol{\gamma}_{m,i}$ for $m = 1, \ldots, M-1$. These constraints ensure that the intensity functions of the latent states are monotone increasing, i.e., $\lambda_i^t(m+1, \mathbf{z}^t, \mathbf{x}^t) > \lambda_i^t(m, \mathbf{z}^t, \mathbf{x}^t)$ for all $\mathbf{x}^t$, $\mathbf{z}^t$, and $m = 1, \ldots, M-1$.

The preceding model describes individual level accelerometer data as a function of evolving covariate and treatment information. The idea of using the framework of hidden Markov model to analyze physical activity data can also be seen in the recent work of [Hua18] and [Kup19]. However, the cat accelerometer data we consider are much more sparse than the human accelerometer data in their work, which motivates a zero-inflation component in our model. Moreover, we can include exogenous, time-varying variables like treatment period, night indicator, and weekend indicator in our model so as to explicitly address their effects on the activity patterns of cats. Our individual level model is a nontrivial extension of the zero-inflated Poisson hidden Markov model in the literature [DB11; OR12], which not only allows for distinct state duration densities but also incorporates covariates and treatment in both transition probabilities and latent state durations.

To draw more general, i.e., population-level, inferences we model individual-level treatment effect parameters as functions of baseline subject information as follows. We assume that $\mathbb{E}\left(\mathbf{b}_{1,m,i}\middle|\mathbf{W}_i=\mathbf{w}_i\right)=\boldsymbol{\Omega}_{m,0}+\boldsymbol{\Omega}_{m,1}\mathbf{w}_i$, where $\boldsymbol{\Omega}_{m,0}\in\mathbb{R}^q$ and $\boldsymbol{\Omega}_{m,1}\in\mathbb{R}^{q\times d}$ are matrices of unknown coefficients encoding the treatment effect on the outcome for $m=0,1,\ldots,M$. Similarly, we assume that $\mathbb{E}\left(\mathbf{c}_{m,i}\middle|\mathbf{W}_i=\mathbf{w}_i\right)=\boldsymbol{\Gamma}_{m,0}+\boldsymbol{\Gamma}_{m,1}\mathbf{w}_i$ and $\mathbb{E}\left(\mathbf{d}_{m,\ell,i}\middle|\mathbf{W}_i=\mathbf{w}_i\right)=\boldsymbol{\Lambda}_{m,\ell,0}+\boldsymbol{\Lambda}_{m,\ell,1}\mathbf{w}_i$ where $\boldsymbol{\Gamma}_{m,0},\boldsymbol{\Lambda}_{m,\ell,0}\in\mathbb{R}^q$ and $\boldsymbol{\Gamma}_{m,1},\boldsymbol{\Lambda}_{m,\ell,1}\in\mathbb{R}^{q\times d}$ encode the effect of treatment on state duration and transition probabilities for $\ell,m=1,\ldots,M$. The idea of aggregating individual-level models to conduct population-level inference has been applied in other contexts including longitudinal and growth curve analyses [SD97; Dig02; VM09].

2.3.1 Estimation

We estimate the subject-specific parameters using maximum-likelihood estimation via the forward-backward recursive representation of the likelihood [Rab89]; the number of latent states, M, is selected using BIC [Sch78]. Subsequently, the population-level parameters are estimated by regressing the subject-specific maximum likelihood estimators on baseline covariates using least squares. This two-stage approach, which will be detailed shortly, is computationally efficient for high-frequency data like those collected in the meloxicam study. Let

$$\boldsymbol{\theta}_i=\left\{\left(\delta_m,\mathbf{d}_{m,k,i},\boldsymbol{\varrho}_{m,k,i},\mathbf{c}_{m,i},\boldsymbol{\eta}_{m,i},b_{0,0,i},b_{0,m,i},\mathbf{b}_{1,0,i},\mathbf{b}_{1,m,i},\boldsymbol{\gamma}_{0,i},\boldsymbol{\gamma}_{m,i}\right)\right\}_{m,k\in\{1,\ldots,M\}}$$

to be the collection of subject-specific parameters for $i=1,\ldots,n$. For each $t=1,\ldots,T$, $m=1,\ldots,M$ and $i=1,\ldots,n$ define the forward-variables $\alpha_i^t(m;\boldsymbol{\theta}_i)=P_{\theta_i}\left(Y_i^1,\ldots,Y_i^t,S_i^t=m\right)$ where P_{θ_i} denotes the distribution indexed by θ_i. We assume there is a finite latent state duration D, and any state durations exceeding D are artificially censored. Then it can be shown [Rab89] that the forward-variables satisfy the following recursions

$$
\begin{aligned}
\alpha_i^1(m;\boldsymbol{\theta}_i) =& \delta_{m,i} r_{m,i}(1) P(Y_i^1|S_i^1 = m),\\
\alpha_i^2(m;\boldsymbol{\theta}_i) =& \delta_{m,i} r_{m,i}(2) \prod_{\tau=1}^{2} P(Y_i^\tau|S_i^\tau = m) +\\
& \sum_{\ell\neq m} \alpha_i^1(\ell;\boldsymbol{\theta}_i) Q_i(\ell,m) r_{m,i}(1) P(Y_i^2|S_i^2 = m),\\
\alpha_i^3(m;\boldsymbol{\theta}_i) =& \delta_{m,i} r_{m,i}(3) \prod_{\tau=1}^{3} P(Y_i^\tau|S_i^\tau = m) +\\
& \sum_{d=1}^{2}\sum_{\ell\neq m} \alpha_i^{3-d}(\ell;\boldsymbol{\theta}_i) Q_i(\ell,m) r_{m,i}(d) \prod_{\tau=4-d}^{3} P(Y_i^\tau|S_i^\tau = m),\\
& \vdots\\
\alpha_i^D(m;\boldsymbol{\theta}_i) =& \delta_{m,i} r_{m,i}(D) \prod_{\tau=1}^{D} P(Y_i^\tau|S_i^\tau = m) +\\
& \sum_{d=1}^{D-1}\sum_{\ell\neq m} \alpha_i^{D-d}(\ell;\boldsymbol{\theta}_i) Q_i(\ell,m) r_{m,i}(d) \prod_{\tau=D+1-d}^{D} P(Y_i^\tau|S_i^\tau = m),
\end{aligned}
$$

where the terms on the right-hand-side of the above equations are dictated by $\boldsymbol{\theta}_i$ through the models posited previously; though we have suppressed this dependence in the notation. For $t = D+1,\ldots,T$ the forward-variables satisfy

$$
\alpha_i^t(m;\boldsymbol{\theta}_i) = \sum_{\ell=1}^{M}\sum_{d=1}^{D}\left[\alpha_i^{t-d}(\ell;\boldsymbol{\theta}_i) Q_i(\ell,m) r_{m,i}(d) \prod_{\tau=t-d+1}^{t} P(Y_i^\tau|S_i^\tau = m)\right].
$$

Thus, the likelihood for $\boldsymbol{\theta}_i$ is $\mathscr{L}_i(\boldsymbol{\theta}_i) = \sum_{m=1}^{M} \alpha_i^T(m;\boldsymbol{\theta}_i)$ and the likelihood for $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_n$ is $\prod_{i=1}^{n}\mathscr{L}_i(\boldsymbol{\theta}_i)$. We construct estimators $\widehat{\boldsymbol{\theta}}_{i,n}$ of $\boldsymbol{\theta}_i$ by maximizing the log-likelihood (see Appendix A for details).

For $m = 0,1,\ldots,M$ let $\widehat{\mathbf{b}}_{1,m,i,T}$ denote the maximum likelihood estimator of $\mathbf{b}_{1,m,i}$. Subsequently, define $\widehat{\boldsymbol{\Omega}}_{m,0,n}, \widehat{\boldsymbol{\Omega}}_{m,1,n} = \arg\min_{\boldsymbol{\Omega}_{m,0}\in\mathbb{R}^q, \boldsymbol{\Omega}_{m,1}\in\mathbb{R}^{q\times d}} \sum_{i=1}^{n} ||\widehat{\boldsymbol{b}}_{1,m,i,T} - \boldsymbol{\Omega}_{m,0} - \boldsymbol{\Omega}_{m,1}\mathbf{W}_i||^2$ to be the two-stage estimator of $\boldsymbol{\Omega}_{m,0}, \boldsymbol{\Omega}_{m,1}$. The estimators $\widehat{\boldsymbol{\Gamma}}_{m,0,n}, \widehat{\boldsymbol{\Gamma}}_{m,1,n}$, $\widehat{\boldsymbol{\Lambda}}_{m,\ell,0,n}$ and $\widehat{\boldsymbol{\Lambda}}_{m,\ell,1,n}$ of $\boldsymbol{\Gamma}_{m,0}, \widehat{\boldsymbol{\Gamma}}_{m,1}$, $\boldsymbol{\Lambda}_{m,\ell,0}$ and $\boldsymbol{\Lambda}_{m,\ell,1}$ are defined analogously.

2.3.2 Theoretical properties

Throughout we use a star superscript, e.g., $\boldsymbol{\theta}_i^*$, to denote the population analogue of the maximum likelihood estimator of a parameter indexing the proposed model; if the model is correctly specified, then under the conditions stated below, these parameters will correspond to the true parameters indexing the generative model. In order to characterize the limiting behavior of the proposed estimators, we make the following assumptions for all $i = 1,\ldots,n$ and $\ell, m \in \{1,\ldots,M\}$.

(A0) The dimension of the subject-specific parameters is $\dim(\boldsymbol{\theta}_i^*) = c$, where c is known; $\boldsymbol{\theta}_i^*$ is an interior point of Θ which is a compact subset of $\mathbb{R}^c$.

(A1) The state duration distribution $r_{m,i}(\cdot)$ has finite support.

(A2) There exists $0 < \sigma_- \leq \sigma_+ < 1$ such that $\sigma_- \leq P_{\boldsymbol{\theta}_i^*}\left(S_i^{t+1} = \ell \middle| S_i^t = m\right) \leq \sigma_+$, for all $t = 1,\ldots,T-1$; and $\sum_{m=1}^M P_{\boldsymbol{\theta}_i^*}\left(Y_i^t = y \middle| S_i^t = m\right) > 0$ for all $y \in \operatorname{supp} Y_i^t$ and $t = 1,\ldots,T$.

(A3) For each $\boldsymbol{\theta} \in \Theta$ the transition kernel indexed by $\boldsymbol{\theta}$ is Harris recurrent and aperiodic. [see AL06; MT12, for details and additional discussion].

(A4) The transition kernel is continuous in $\boldsymbol{\theta}$ in an open neighborhood of $\boldsymbol{\theta}_i^*$.

(A5) The hidden semi-Markov model is identifiable up to label-switching [see All09; Gas13, for discussion of label-switching].

These assumptions are standard in latent-state models [Ler92; Bic98; LGM00; DM01; Dou04]. Assumption (A0) avoids non-regularity occurring at boundary points; the assumption of a fixed dimension could be relaxed, for example, to the assumption that one has a strongly consistent estimator of c. Assumption (A1) is common in semi-Markov models and simplifies asymptotic arguments; if the finite support condition does not hold, one can use a nested sequence of finite approximations at the expense of more delicate asymptotic arguments [BL09; TL11]. Assumption (A4) is a regularity condition that avoids non-standard asymptotic behavior associated with non-smooth functionals [VDV91; PP13]. Assumptions (A2), (A3), (A5) ensure that the model is well-defined; we show that (A5) holds for the zero-inflated Poisson semi-Markov model in Appendix A.

This result is anticipated by previous work on hidden Markov models [Ler92; Bic98; LGM00; DM01; Dou04].

Lemma 2.3.1. *Under (A0) - (A5), the MLE for the subject-specific hidden semi-Markov model in the first stage is strongly consistent, i.e.,* $\hat{\boldsymbol{\theta}}_{i,T} \to \boldsymbol{\theta}_i^*$ *almost surely as* $T \to \infty$ *for all* $i = 1,\ldots,n$.

Define $\boldsymbol{\Omega}_m^* \triangleq [\boldsymbol{\Omega}_{m,0}^*, \boldsymbol{\Omega}_{m,1}^*]$ and the corresponding estimator $\widehat{\boldsymbol{\Omega}}_{m,n} \triangleq [\widehat{\boldsymbol{\Omega}}_{m,0,n}, \widehat{\boldsymbol{\Omega}}_{m,1,n}]$. Let, $\boldsymbol{\Gamma}_m^*$, $\widehat{\boldsymbol{\Gamma}}_{m,n}$, $\boldsymbol{\Lambda}_{m,\ell}^*$, and $\widehat{\boldsymbol{\Lambda}}_{m,\ell,n}$ be defined analogously. Further, define the augmented design matrix $\tilde{\mathbf{W}} \in \mathbb{R}^{n\times(d+1)}$ whose i^{th} row is $\tilde{\mathbf{W}}_i \triangleq [1, \mathbf{W}_i^T] \in \mathbb{R}^{d+1}$. Then the second stage regression models can be equivalently written as $\mathbf{vec}(\widehat{b}_{1,m,T}) = (\boldsymbol{I} \otimes \tilde{\boldsymbol{W}})\mathbf{vec}(\boldsymbol{\Omega}_m) + \boldsymbol{\upsilon}_m$, $\mathbf{vec}(\widehat{c}_{m,T}) = (\boldsymbol{I} \otimes \tilde{\boldsymbol{W}})\mathbf{vec}(\boldsymbol{\Gamma}_m) + \boldsymbol{\eta}_m$, $\mathbf{vec}(\widehat{d}_{m,\ell,T}) = (\boldsymbol{I} \otimes \tilde{\boldsymbol{W}})\mathbf{vec}(\boldsymbol{\Lambda}_{m,\ell}) + \boldsymbol{\xi}_{m,\ell}$, where **vec** is the vectorization operator, and $\boldsymbol{\upsilon}_m, \boldsymbol{\eta}_m, \boldsymbol{\xi}_{m,\ell}$ are independent mean zero residuals. Define $\tilde{\boldsymbol{B}} \triangleq \boldsymbol{I} \otimes \tilde{\boldsymbol{W}} \in \mathbb{R}^{qn\times q(d+1)}$ whose i^{th} row is $\tilde{\boldsymbol{B}}_i$.

Theorem 2.3.1. *Assume (A0) - (A5), and further that $\boldsymbol{H} = n^{-1} \lim_{n\to\infty} \tilde{\boldsymbol{W}}^T \tilde{\boldsymbol{W}}$ is a positive definite matrix. Then, provided $T, n \to \infty$ with $n/T \to 0$, each of the following converges in distribution to a Gaussian distribution with mean zero and identity covariance:*

$$\left(\left[\frac{\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}}{n}\right]^{-1}\left[\frac{1}{n}\tilde{\boldsymbol{B}}^T\mathbf{diag}\left\{[\mathbf{vec}(\widehat{\boldsymbol{b}}_{1,m,T})_i - \tilde{\boldsymbol{B}}_i^T\mathbf{vec}(\widehat{\boldsymbol{\Omega}}_{m,n})]^2\right\}\tilde{\boldsymbol{B}}\right]\left[\frac{\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}}{n}\right]^{-1}\right)^{-1/2} \times$$
$$\sqrt{n}\left\{\mathbf{vec}(\widehat{\boldsymbol{\Omega}}_{m,n}) - \mathbf{vec}(\boldsymbol{\Omega}_m^*)\right\}$$
$$\left(\left[\frac{\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}}{n}\right]^{-1}\left[\frac{1}{n}\tilde{\boldsymbol{B}}^T\mathbf{diag}\left\{[\mathbf{vec}(\widehat{\boldsymbol{c}}_{m,T})_i - \tilde{\boldsymbol{B}}_i^T\mathbf{vec}(\widehat{\boldsymbol{\Gamma}}_{m,n})]^2\right\}\tilde{\boldsymbol{B}}\right]\left[\frac{\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}}{n}\right]^{-1}\right)^{-1/2} \times$$
$$\sqrt{n}\left\{\mathbf{vec}(\widehat{\boldsymbol{\Gamma}}_{m,n}) - \mathbf{vec}(\boldsymbol{\Gamma}_m^*)\right\}$$
$$\left(\left[\frac{\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}}{n}\right]^{-1}\left[\frac{1}{n}\tilde{\boldsymbol{B}}^T\mathbf{diag}\left\{[\mathbf{vec}(\widehat{\boldsymbol{d}}_{m,\ell,T})_i - \tilde{\boldsymbol{B}}_i^T\mathbf{vec}(\widehat{\boldsymbol{\Lambda}}_{m,\ell,n})]^2\right\}\tilde{\boldsymbol{B}}\right]\left[\frac{\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}}{n}\right]^{-1}\right)^{-1/2} \times$$
$$\sqrt{n}\left\{\mathbf{vec}(\widehat{\boldsymbol{\Lambda}}_{m,\ell,n}) - \mathbf{vec}(\boldsymbol{\Lambda}_{m,\ell}^*)\right\},$$

for all $m, \ell \in \{1, \ldots, M\}$.

In the preceding theorem, the requirement that $T/n \to 0$ is natural in applications using high-frequency accelerometer data where the number of observations per subject can be several orders of magnitude larger than the number of subjects.

2.4 Simulation experiments

We study the finite sample performance of the proposed two-stage estimator using a suite of simulation experiments. Across these experiments, we consider three different pathways for the treatment effect: (Scenario I) a change in the state-dependent means and the proportion of zeros; (Scenario II) a change in the latent state durations; and (Scenario III) a change in the state-transition probabilities. In each scenario, we simulate minute-by-minute activity counts during a two week period for $n = 10$ and $n = 20$ subjects. Each simulated trajectory is divided into an initial one-week baseline phase followed by a one-week treatment period; thus, each patient has 10,080 observations per period. The activity counts are generated using a three state zero-inflated Poisson hidden semi-Markov model, with the maximum latent state duration set to be 20. The initial state probabilities are distributed as follows: let

$U_1, U_2 \sim_{i.i.d.}$ Uniform(0,1) and set $\delta_1 = \min(U_1, U_2)$, $\delta_2 = \max(U_1, U_2) - \delta_1$, and $\delta_3 = 1 - \delta_1 - \delta_2$. We generate subject covariates so that half of the subjects are male, and set the baseline Poisson means for the male subjects to be 5% less than the female subjects while the baseline zero proportion is 5% larger than the female subjects. Furthermore, for all subjects (regardless of sex) the Poisson means are 5% less at night than during the day and the zero proportions are 5% larger at night than during the day. These proportions are consistent with the accelerometer data collected in the study of the treatment effect of meloxicam.

In Scenario I, treatment increases the log mean activity in all the latent states by 10% and decreases the log odds of zero in state 1 by 10%, so that

$$\begin{aligned} \log\left\{\frac{p_i^t(\mathbf{z}^t, \mathbf{x}^t)}{1 - p_i^t(\mathbf{z}^t, \mathbf{x}^t)}\right\} &= b_{0,0,i} + 0.05 \times \mathbf{I}\{\text{Male}\}_i - 0.1 \times \mathbf{I}\{\text{Trt}\}_i^t + 0.05 \times \mathbf{I}\{\text{Night}\}_i^t, \\ \log\left\{\lambda_i^t(m, \mathbf{z}^t, \mathbf{x}^t)\right\} &= b_{0,m,i} - 0.05 \times \mathbf{I}\{\text{Male}\}_i + 0.1 \times \mathbf{I}\{\text{Trt}\}_i^t - 0.05 \times \mathbf{I}\{\text{Night}\}_i^t, \end{aligned}$$

for $m = 1, 2, 3$, where $b_{0,0,i} \overset{iid}{\sim} N(0, 0.1^2)$, $b_{0,1,i} \overset{iid}{\sim} N\left\{\log(50), 0.1^2\right\}$, $b_{0,2,i} \overset{iid}{\sim} N\left\{\log(300), 0.1^2\right\}$, and $b_{0,3,i} \overset{iid}{\sim} N\left\{\log(700), 0.1^2\right\}$. Let $N_{[a,b]}(\mu, \sigma^2)$ denote a normal distribution with mean μ and variance σ^2 truncated to $[a, b]$. In this scenario, we assume that both the transition probabilities and the latent state durations are unaffected by the treatment and that $\mathbf{Q}_i(1,2) \overset{iid}{\sim} N_{[0,1]}(0.5, 0.01)$, $\mathbf{Q}_i(2,1) \overset{iid}{\sim} N_{[0,1]}(0.7, 0.01)$, $\mathbf{Q}_i(3,1) \overset{iid}{\sim} N_{[0,1]}(0.7, 0.01)$, so that the transition matrix for subject i is

$$\begin{bmatrix} 0 & \mathbf{Q}_i(1,2) & 1 - \mathbf{Q}_i(1,2) \\ \mathbf{Q}_i(2,1) & 0 & 1 - \mathbf{Q}_i(2,1) \\ \mathbf{Q}_i(3,1) & 1 - \mathbf{Q}_i(3,1) & 0 \end{bmatrix}.$$

This setup mimics the real accelerometer data where state 1 (lowest activity level) dominates the other latent states. We do not allow transition back to the same state as the latent state durations modeled separatetly; in Scenario 1, the latent state durations are uniformly distributed on $\{1, 2, \ldots, 20\}$ for each state.

In Scenario II, we assume that the treatment increases the log mean activity in state 3 by 10% but decreases the log mean activity in state 1 and 2 by 10%. Further, the treatment increases the log odds of zero in state 1 by 10%, such that

$$
\begin{aligned}
\log\left\{\frac{p_i^t(\mathbf{z}^t,\mathbf{x}^t)}{1-p_i^t(\mathbf{z}^t,\mathbf{x}^t)}\right\} &= b_{0,0,i}+0.05\times\mathbf{I}\{\text{Male}\}_i+0.1\times\mathbf{I}\{\text{Trt}\}_i^t+0.05\times\mathbf{I}\{\text{Night}\}_i^t,\\
\log\left\{\lambda_i^t(1,\mathbf{z}^t,\mathbf{x}^t)\right\} &= b_{0,1,i}-0.05\times\mathbf{I}\{\text{Male}\}_i-0.1\times\mathbf{I}\{\text{Trt}\}_i^t-0.05\times\mathbf{I}\{\text{Night}\}_i^t,\\
\log\left\{\lambda_i^t(2,\mathbf{z}^t,\mathbf{x}^t)\right\} &= b_{0,2,i}-0.05\times\mathbf{I}\{\text{Male}\}_i-0.1\times\mathbf{I}\{\text{Trt}\}_i^t-0.05\times\mathbf{I}\{\text{Night}\}_i^t,\\
\log\left\{\lambda_i^t(3,\mathbf{z}^t,\mathbf{x}^t)\right\} &= b_{0,3,i}-0.05\times\mathbf{I}\{\text{Male}\}_i+0.1\times\mathbf{I}\{\text{Trt}\}_i^t-0.05\times\mathbf{I}\{\text{Night}\}_i^t,
\end{aligned}
$$

where $b_{0,0,i}\overset{iid}{\sim}N(0,0.01)$, $b_{0,1,i}\overset{iid}{\sim}N(\log(50),0.01)$, $b_{0,2,i}\overset{iid}{\sim}N(\log(300),0.01)$, $b_{0,3,i}\overset{iid}{\sim}N(\log(700),0.01)$. In this scenario, treatment will increase activity in the highest level state but make the subject less agitated in lower level activity states. In this scenario, both the transition probabilities and the latent state durations are unaffected by the treatment and match the settings of Scenario I.

In Scenario III, we assume that treatment accelerates the duration time by 10% in each latent state, so that subjects are more likely to actively switch between states when they are on treatment, i.e.,

$$
r_{m,i}\left(v;\mathbf{x}^t,\mathbf{z}^t\right)=\exp\left(0.1\times\mathbf{I}\{\text{Trt}\}_i^t\right)\int_v^{v+1} f_{m,i}\left\{\exp\left(0.1\times\mathbf{I}\{\text{Trt}\}_i^t\right)u\right\}du,
$$

for $m=1,2,3$, where $f_{1,i}$, $f_{2,i}$, $f_{3,i}$ are the exponential densities with rate parameters $\xi_{1,i}$, $\xi_{2,i}$, $\xi_{3,i}$, such that $\xi_{1,i}\overset{iid}{\sim}N(5,0.1^2)$, $\xi_{2,i}\overset{iid}{\sim}N(3,0.1^2)$, and $\xi_{3,i}\overset{iid}{\sim}N(2,0.1^2)$. This setup also mimics the study of the treatment effect of the meloxicam in that the expected duration in state 1 is longer than in other states. We assume that the treatment does not affect the state-dependent Poisson means and zero proportion so that

$$
\begin{aligned}
\log\left\{\frac{p_i^t(\mathbf{z}^t,\mathbf{x}^t)}{1-p_i^t(\mathbf{z}^t,\mathbf{x}^t)}\right\} &= b_{0,0,i}+0.05\times\mathbf{I}\{\text{Male}\}_i+0.05\times\mathbf{I}\{\text{Night}\}_i^t\\
\log\left\{\lambda_i^t(m,\mathbf{z}^t,\mathbf{x}^t)\right\} &= b_{0,m,i}-0.05\times\mathbf{I}\{\text{Male}\}_i-0.05\times\mathbf{I}\{\text{Night}\}_i^t,
\end{aligned}
$$

for all $m=1,2,3$, $b_{0,0,i}\overset{iid}{\sim}N(0,0.1^2)$, $b_{0,1,i}\overset{iid}{\sim}N\left\{\log(50),0.1^2\right\}$, $b_{0,2,i}\overset{iid}{\sim}N\left\{\log(300),0.1^2\right\}$, $b_{0,3,i}\overset{iid}{\sim}N\left\{\log(700),0.01\right\}$. The transition probabilities are the same as in Scenario I.

In Scenario IV, we assume that treatment affects the state transition probabilities by increasing the odds of moving to the next most active state by 10% so that

$$\log\left\{\frac{Q_i(1,3;\mathbf{x}^t,\mathbf{z}^t)}{Q_i(1,2;\mathbf{x}^t,\mathbf{z}^t)}\right\} = \eta_{1,i} + 0.1 \times \mathbf{I}\{\text{Trt}\}_i^t,$$
$$\log\left\{\frac{Q_i(2,3;\mathbf{x}^t,\mathbf{z}^t)}{Q_i(2,1;\mathbf{x}^t,\mathbf{z}^t)}\right\} = \eta_{2,i} + 0.1 \times \mathbf{I}\{\text{Trt}\}_i^t,$$
$$\log\left\{\frac{Q_i(3,2;\mathbf{x}^t,\mathbf{z}^t)}{Q_i(3,1;\mathbf{x}^t,\mathbf{z}^t)}\right\} = \eta_{3,i} + 0.1 \times \mathbf{I}\{\text{Trt}\}_i^t,$$

where $\eta_{1,i} \overset{iid}{\sim} N(0,0.1^2)$, $\eta_{2,i} \overset{iid}{\sim} N(-0.5,0.1^2)$, $\eta_{1,i} \overset{iid}{\sim} N(-0.5,0.1^2)$. The latent state duration are $r_{m,i}(v) = \int_v^{v+1} f_{m,i}\{u\}\,du$, for $m = 1,2,3$, where $f_{1,i}$, $f_{2,i}$, $f_{3,i}$ are the exponential densities with rate parameters $\xi_{1,i}$, $\xi_{2,i}$, $\xi_{3,i}$ respectively, such that $\xi_{1,i} \overset{iid}{\sim} N(5,0.1^2)$, $\xi_{2,i} \overset{iid}{\sim} N(3,0.1^2)$, $\xi_{3,i} \overset{iid}{\sim} N(2,0.1^2)$. The state-dependent Poisson means and zero proportion are as in Scenario III.

We compare the proposed estimator with the baseline treatment effect estimator constructed using a generalized estimating equation (GEE). This is equivalent to the repeated-measures ANOVA model used in Gruen et al. [Gru15] and Gruen et al. [Gru16] as the data are complete and balanced. In the GEE model, denote the new response variable in the model by $\tilde{Y}_i^\tau$, which is the average activity count for subject i on day $\tau = 1,\ldots,14$. Define $\mathbf{¥}_i \triangleq [\tilde{Y}_i^1,\ldots,\tilde{Y}_i^{14}]$. Let $E(\mathbf{¥}_i|\tilde{\mathbf{X}}_i) = \tilde{\mathbf{X}}_i\boldsymbol{\beta}$ and $\text{var}(\mathbf{¥}_i|\tilde{\mathbf{X}}_i) = \mathbf{V}_i$, where $\tilde{\mathbf{X}}_i \in \mathbb{R}^{14\times 3}$ whose τ^{th} row is $[1,\mathbf{I}\{\text{Trt}\}_i^\tau,\mathbf{I}\{\text{Male}\}_i]$; $\boldsymbol{\beta} = [\beta_0,\beta_1,\beta_2]$ so that β_1 represents the treatment effect; the working covariance $\mathbf{V}_i$ is assumed to be compound symmetric.

Table 4.1 shows the mean and standard error for the estimated treatment effect estimators in the proposed method using 500 Monte Carlo replications. The estimators are approximately unbiased in all four scenarios and the standard error is small even with a moderate number of samples owing to within-subject pooling across observations. A summary of the simulation runtimes is included in Appendix A. Table 2.2 shows the mean and standard error for the estimated treatment effect in the baseline GEE model. The GEE method manages to detect the treatment effect in Scenarios I and IV, where there is an increase in the marginal mean activity counts under treatment; however, it fails to detect the nonzero conditional treatment effect in different latent states using an $\alpha = 0.05$ level test in Scenario II and III because there is only a small difference in the marginal mean activity counts.

Table 2.1 Mean (standard error) of the treatment effect estimators from the proposed hidden semi-markov model in Scenarios I, II, III, and IV for each latent state based on 500 Monte Carlo simulations.

Parameter	True effect	n=10 Estimated effect (s.e.)	n=20 Estimated effect (s.e.)
Scenario I			
State 1: $\widehat{\Omega}_{0,0}$	-0.1	-.0997 (.0159)	-.0994 (.0113)
State 1: $\widehat{\Omega}_{1,0}$	0.1	.0999 (.0015)	.0999 (.0011)
State 2: $\widehat{\Omega}_{2,0}$	0.1	.1000 (.0007)	.1000 (.0005)
State 3: $\widehat{\Omega}_{3,0}$	0.1	.1000 (.0007)	.1000 (.0005)
Scenario II			
State 1: $\widehat{\Omega}_{0,0}$	0.1	.0994 (.0159)	.1002 (.0116)
State 1: $\widehat{\Omega}_{1,0}$	-0.1	-.1000 (.0017)	-.1001 (.0012)
State 2: $\widehat{\Omega}_{2,0}$	-0.1	-.1001 (.0008)	-.1000 (.0006)
State 3: $\widehat{\Omega}_{3,0}$	0.1	.0999 (.0007)	.1000 (.0005)
Scenario III			
State 1: $\widehat{\Gamma}_{1,0}$	0.1	.0931 (.0177)	.0931 (.0131)
State 2: $\widehat{\Gamma}_{2,0}$	0.1	.0931 (.0216)	.0920 (.0158)
State 3: $\widehat{\Gamma}_{3,0}$	0.1	.0935 (.0222)	.0952 (.0167)
Scenario IV			
State 1: $\widehat{\Lambda}_{1,3,0}$	0.1	.1006 (.0414)	.1022 (.0290)
State 2: $\widehat{\Lambda}_{2,3,0}$	0.1	.0983 (.0486)	.1006 (.0346)
State 3: $\widehat{\Lambda}_{3,2,0}$	0.1	.0995 (.0447)	.1001 (.0321)

Table 2.2 Mean (standard error) of the treatment effect estimators from the baseline GEE model for Scenario I, II, III, and IV via 500 Monte Carlo simulations.

Scenario	n=10	n=20
	Estimated treatment effect (s.e.)	Estimated treatment effect (s.e.)
I	19.77 (1.86)	19.57 (1.26)
II	0.46 (1.94)	0.54 (1.24)
III	1.21 (1.91)	1.16 (1.33)
IV	6.02 (2.08)	5.88 (1.48)

2.5 Results

We applied the proposed zero-inflated hidden semi-Markov model to the longitudinal accelerometer data from the study of the treatment effect of meloxicam. The primary hypothesis is that meloxicam reduces OA-associated pain which will manifest through increased activity during the treatment period relative to the placebo period owing to increased intensity of movements in active states.

In the first stage model we included as covariates: treatment status, weekend (yes/no), and night (yes/no). We choose the number of latent states to be six using minimum BIC as our selection criterion. In the second stage model, we included as covariates: sex, Body Condition Score (BCS), age, total OA score, and treatment sequence. Continuous covariates were centered before fitting the model.

Table 2.3 displays the estimated treatment effects in terms of mean activity counts within each latent state. There are significant effects ($p < 0.05$) in latent states 4-6 suggesting that cats treated with meloxicam may have increased mean activity while engaged in more intensive activities (e.g., running, jumping, etc.). The effect size of the treatment on those more intensive activities ranges from 7.28% to 9.14%. Table 2.4 shows the mean proportion of time spent in each latent state estimated using the Viterbi algorithm [Vit67]. It can be seen that the proportion of time spent in each activity is similar across the treatment and placebo periods. Figure 2.3 shows a QQ-plot of the estimated quantiles using the fitted model and the observed quantiles. It can be seen that they follow each other closely and that there are no indications of serious lack-of-fit.

Table 2.3 Summary of treatment effects (meloxicam vs. placebo) in each latent state. The primary hypothesis is that meloxicam reduces OA-associated pain which will manifest through increased activity during the treatment period relative to the placebo period.

Parameter	Estimate	Robust s.e.	P-value
State 1 (zero)	-0.0299	0.0433	0.4899
State 1 (mean)	0.0222	0.0181	0.2187
State 2 (mean)	0.0444	0.0311	0.1538
State 3 (mean)	0.0524	0.0318	0.0990
State 4 (mean)	0.0914	0.0351	0.0091
State 5 (mean)	0.0819	0.0235	0.0005
State 6 (mean)	0.0728	0.0260	0.0051

Table 2.4 Mean proportion of activities in each decoded latent state estimated by Viterbi algorithm [Vit67] in the fitted zero-inflated Poisson hidden semi-Markov model.

State	meloxicam	Placebo
1	83.6%	84.5%
2	5.5%	5.3%
3	4.5%	4.3%
4	3.4%	3.3%
5	2.1%	1.9%
6	0.9%	0.8%

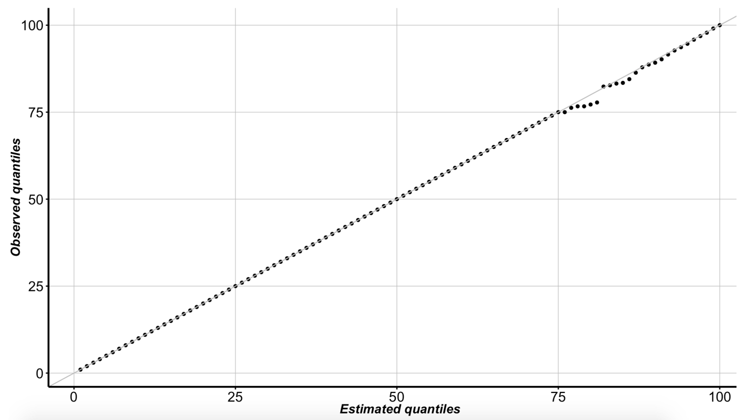

Figure 2.3 Estimated quantiles of activity counts based on the fitted zero-inflated Poisson hidden semi-Markov model versus observed quantiles.

2.6 Discussion

We proposed a zero-inflated Poisson hidden semi-Markov model for activity data captured by accelerometer readings in a naturally occurring model of OA-associated pain. The proposed model can be used to characterize the impact of treatment on activity including changes in activity duration, intensity, and frequency. In the context of OA-associated pain, understanding such impacts is critical to fully understanding the effects of treatment on patient mobility and therefore quality of life. In the study of the treatment effect of meloxicam, our model found that meloxicam significantly improves the mobility of cats in more intensive activities, while the activity patterns did not change in less intensive activities.

CHAPTER

3

HIERARCHICAL CONTINUOUS TIME HIDDEN MARKOV MODEL, WITH APPLICATION IN ZERO-INFLATED ACCELEROMETER DATA

3.1 Introduction

The development of the wearable technology has given rise to a variety of sensing devices and modalities. Some of these devices, e.g., Fitbit or Apple Watch, can be worn continuously and thereby produce huge volumes of high-frequency human activity data. Because these data present little burden on the wearer to collect and provide rich information on the *in situ* behavior of the wearer, they have tremendous potential to inform decision making in healthcare. Examples of remote sensing data in healthcare include elder care, remote monitoring of chronic disease, and addition management [Bar05; Mar07; Han12]. Accelerometers are among the most commonly used and most widely studied types of wearable devices, they have been used both in randomized clinical trials to evaluate treatment effect on activity-related impairment [Nap10; Kan18] and in observational studies to characterize activity patterns in a free-living environment [Mor06; Han12; Xia14]. However, despite rapidly growing interest and investment in wearable devices for the study of human activity data, a general and extensible class of models for analysis of the resulting data is lacking.

We propose a continuous-time hidden Markov model for the modeling human accelerometer data that aligns with scientific (conceptual) models of human activity data; in the proposed model, latent states correspond to latent (unobserved) activities, e.g., resting, running, jumping, etc., that are shared across the population yet the accelerometer signatures within these activities are allowed to vary across subjects. Furthemore, differences across subgroups, e.g., defined by sex, age, or the presence-absence of a comorbid condition, can be identified by aggregating individual-level effects across these groups. This work is motivated in part by the physical activity data set from the 2003 - 2004 National Health and Nutrition Examination Survey (NHANES). In this study, human activity patterns were measured at one-minute intervals for up to seven days using the ActiGraph Model 7164 accelerometer [Tro08; Met08; Sch15]. Activity for each minute was recorded as an integer-valued intensity-level commonly referred to as an activity count. In the study, subjects were instructed to remove the device during sleep or while washing (to keep it dry). Therefore, the observed data comprise high-frequency, integer-valued activity counts for each subject with intervals of missing values corresponding to when the device was removed.

The goal of paper is to use the observed data to characterize activity patterns of each subject, subjects within pre-defined subgroups, and the population as a whole. This is important because the estimated physical activity model can potentially serve the following three purposes. On the subject level, the estimated activity model can be used both for the prediction of future activities and the imputation of missing activity readings. On the subgroup level, the estimated activity patterns provide useful insights into clustering people based on their activity profiles. On the population level, public policies can be designed based on the estimated activity model so as to encourage everyday exercise and healthy life style.

Prior work on modeling activity counts has focused on aggregation and other smoothing techniques. One common approach is to average the activity counts over time for each subject and then compare the group means using two-sample t-tests [Tro08], analysis of covariance [Han12], or linear mixed effects models [Cra04]. However, in these approaches, averaging focuses on overall activity levels and may obscure trends in activity type, activity duration, and transitions between activities. Another approach is to use functional data analysis methods wherein the integer activity counts are first log-transformed to fix the right-skewedness in the distribution and the transformed activity modeled as a function of time of day and other covariates [Mor06; Xia14; Gru17]. These approaches are best-suited for the identification of smooth, cyclical patterns in the data whereas the observed accelerometer data are characterized by abrupt (i.e., non-smooth) changes in activity levels.

Discrete-time hidden Markov models are another common approach to the analysis of mobility data measured by wearable devices [He07; Nic11; RC14; Wit14]. In these models, activity is partitioned into different latent behavioral states and the observed activity count is dependent on the unobserved latent activity. The latent states evolve according to a discrete-time Markov process and a primary goal is the correct classification of the latent activity. To construct and validate these models requires training data that are labeled by latent activity. However, the NHANES data, like many accelerometer studies, are not labeled by activity. Furthermore, our goal is to identify the dynamics of a patients evolution through these latent activities including activity duration, activity intensity, and transitions between activities. Discrete-time hidden Markov models have been used to model latent health states and subsequently conduct inference for activity patterns within each state [Sco05; Alt07; Shi10], but the time scales in these applications are rather coarse (daily or weekly). By contrast, the physical activity in the NHANES data is measured for each minute; this results in a much larger data volume and the ability to provide are more complete picture of activity dynamics.

A technical limitation of the discrete-time approach, is that it assumes that the observations are equally spaced in time. Continuous-time hidden Markov models (CTHMM) have been used to analyze the irregularly-sampled temporal measurements [Nod12; Wan14; Liu15]. The flexibility of the CTHMM comes at the expense of increased computational cost, which makes it infeasible for large datasets without modification. [Liu15] developed an efficient learning algorithm for parameter estimation in the CTHMMs. However, this algorithm is only suitable for either the completely pooled or unpooled cases wherein all subjects are assumed to be either completely homogeneous so that they share the same parameters, or completely heterogeneous so that all parameters are subject-specific. Moreover, the algorithm cannot estimate the effects of subject covariates and environmental factors on activity counts.

We propose to model minute-by-minute accelerometer data using a hierarchical continuous-time hidden Markov model (HCTHMM). This model is aligned with scientific models of activity as the latent states represent different types of unobserved physical activities. The continuous-time Markov process for the latent states evolution avoids having to perform imputation for missing yet allows for the possibility that the temporal measurements are irregularly spaced. Furthermore, the proposed model can incorporate both baseline subject covariates and time-varying environmental factors. The proposed model is hierarchical in that it is parameterized by: (i) subject-specific parameters to account for variability between subjects; (ii) subgroup-specific parameters parameters to account for similarity in activity patterns within groups; (iii) population parameters that are common across all subjects. This specification allows us to pool information on some parameters while retaining between-group and between-subject variability. We proposed an estimator of these parameters that is based on consensus optimization using the alternating direction method of multipliers (ADMM). There is a vast literature on the convergence properties of ADMM [Boy11; Shi14; HL17] which can be readily ported to the proposed algorithm. Finally, we use the nonparametric bootstrap [Efr92] to estimate the sampling distributions of parameter estimators and to conduct statistical inference.

3.2 Model Framework

We assume that the observed data are of the form $\{\mathbf{W}_i, Y_i(\mathbf{T}_i), \mathbf{X}_i(\mathbf{T}_i)\}_{i=1}^n$, which comprise n independent copies of the trajectory $\{\mathbf{W}, Y(\mathbf{T}), \mathbf{X}(\mathbf{T})\}$, where: $\mathbf{W} \in \mathbb{R}^p$ are baseline subject characteristics; $Y(\mathbf{T}) = \{Y(T_1), \ldots, Y(T_K)\}$ are the non-negative integer activity counts at times $\mathbf{T} = (T_1, \ldots, T_K) \in [0,1]^K$; and $\mathbf{X}(\mathbf{T}) = \{\mathbf{X}(T_1), \ldots, \mathbf{X}(T_K)\}$ are concurrent environmental factors such that $\mathbf{X}(\cdot) \in \mathbb{R}^q$. Both $\mathbf{T}$ and K are treated as random variables as the number and timing of observations vary across subjects. We model the evolution of the observed data using a hierarchical continuous-time hidden Markov model (HCTHMM), which we will develop over the remainder of this section.

Let $S_i(t) \in \{1, \ldots, M\}$ denote the unobserved latent state for subject $i = 1, \ldots, n$ at time $t \in [0,1]$. The latent state evolves according to a Markov process indexed by: (i) an initial state distribution $\pi_i(m) \triangleq P\{S_i(0) = m)\}$ for $m = 1, \ldots, M$ such that $\sum_{m=1}^M \pi_i(m) = 1$; (ii) a transition rate matrix $\mathbf{Q}_i = \{q_i(m,\ell)\}_{m,\ell=1,\ldots,M}$ such that $q_i(m,m) = -\sum_{\ell \neq m} q_i(m,\ell)$. The transition rate matrix, also known as the infinitesimal generator matrix, describes the rate of movements between states in a continuous-time Markov chain [Pyk61a; Pyk61b; Alb62]; the transition probabilities are derived from the transition rates through a matrix exponential operation such that for $k = 1, \ldots, K-1$ and $t > u$, $P_i^{t-u}(m,\ell) \triangleq P\{S_i(t) = \ell | S_i(u) = m\} = \{e^{(t-u)\mathbf{Q}_i}\}_{m,\ell}$.

We assume that the conditional distribution of the activity counts is homogeneous in time given the current latent state and environmental factors (to streamline notation, we include baseline characteristics in the time-varying environmental factors). For $i = 1, \ldots, n$ and $m = 1, \ldots, M$ define $g_{i,m}(y; \mathbf{x}) \triangleq P\{Y_i(t) = y | S_i(t) = m, \mathbf{X}_i(t) = \mathbf{x}\}$. Because longitudinal activity count data are zero-inflated, we set $g_{i,1}(y; \mathbf{x})$ to be the probability mass function for a zero-inflated Poisson distribution with structural zero proportion δ_i and mean $\lambda_{i,1}$ for state 1 such that

$$g_{i,1}(y;\mathbf{x}) = \delta_i(\mathbf{x})1_{y=0} + \{1-\delta_i(\mathbf{x})\}\frac{\left\{\lambda_{i,1}(\mathbf{x})\right\}^y \exp\left\{\lambda_{i,1}(\mathbf{x})\right\}}{y!}.$$

For $m = 2,\ldots,M$ we model the activity counts using a Poisson regression model so that

$$g_{i,m}(y;\mathbf{x}) = \frac{\left\{\lambda_{i,m}(\mathbf{x})\right\}^y \exp\left\{\lambda_{i,m}(\mathbf{x})\right\}}{y!}.$$

For each subject i, and latent state m, we assume that functions $\delta_i(\mathbf{x})$ and $\lambda_{i,m}(\mathbf{x})$are of the form

$$\begin{aligned}
\log\left\{\frac{\delta_i(\mathbf{x})}{1-\delta_i(\mathbf{x})}\right\} &= b_{i,0,0} + \mathbf{x}'\mathbf{b}_{i,0,1}, \\
\log\left\{\lambda_{i,m}(\mathbf{x})\right\} &= b_{i,m,0} + \mathbf{x}'\mathbf{b}_{i,m,1},
\end{aligned}$$

where $b_{i,0,0},\ldots,b_{i,M,0}$ and $\mathbf{b}_{i,0,1},\ldots,\mathbf{b}_{i,M,1}$ are unknown coefficients.

In the foregoing model description, all parameters are subject-specific so that each subject's trajectory can be modeled separately. However, in the HCTHMM, some of the parameters are shared among pre-defined subgroups of the subjects. We assume that subjects are partitioned into J such subgroups based on their baseline characteristics $\mathbf{W}$. For example, these groups might be determined by age and sex. Subjects within the same group are though to behave more similarly to each other than across groups. Let $G_i \in \{1,\ldots,J\}$ be the subgroup to which subject i belongs, and let n_j denote the number of subjects in group $j = 1,\ldots,J$.

The HCTHMM is a flexible multilevel model in that it allows for three levels of of parameters: (i) subject-specific, (ii) subgroup-specific, (iii) population-level. For example, one might let the intercepts in the generalized linear models for state-dependent parameters be subject-specific to account for the between-subject variability; let the initial state probabilities and the transition rate parameters depend on group-membership, i.e. $\pi_{i_1} = \pi_{i_2}$ and $\mathbf{Q}_{i_1} = \mathbf{Q}_{i_2}$ for all i_1, i_2 such that $G_{i_1} = G_{i_2}$; and let the slope parameters in the generalized linear models for state-dependent parameters be common across all subjects.

If all the observed time points are equally spaced and all the parameters are subject-specific, the HCTHMM reduces to the subject-specific zero-inflated Poisson hidden Markov model. If there are no covariates and all parameters are common for all subjects, the HCTHMM reduces to a zero-inflated variant of the continuous-time hidden Markov model [Liu15]. The extension from the previous models to the HCTHMM better matches the scientific goals associated with analyzing the NHANES data but also requires new methods for estimation. Because of the hierarchical structure in the parameters, joint parameter estimation is no longer embarrassingly parallelizable as it would be in the case of its completely pooled or unpooled counterparts.

3.3 Parameter Estimation

3.3.1 Forward-Backward Algorithm

For subject $i = 1, \ldots, n$, let $\mathbf{a}_i \in \mathbb{R}^{M-1}$ be the $M-1$ free parameters in the initial probabilities π_i, and let $\mathbf{c}_i \in \mathbb{R}^{M(M-1)}$ be the $M(M-1)$ free parameters in the transition matrix $\mathbf{Q}_i$. To simplify notation, write $\mathbf{b}_{i,0} \triangleq [b_{i,0,0}, \ldots, b_{i,M,0}] \in \mathbb{R}^{M+1}$ and $\mathbf{b}_{i,1} \triangleq [\mathbf{b}'_{i,0,1}, \ldots, \mathbf{b}'_{i,M,1}] \in \mathbb{R}^{q(M+1)}$ to denote the parameters indexing the generalized linear models for the activity counts in each state. Define the entire vector of parameters for subject i to be $\boldsymbol{\theta}_i \triangleq [\mathbf{a}'_i, \mathbf{c}'_i, \mathbf{b}'_{i,0}, \mathbf{b}'_{i,1}] \in \mathbb{R}^{(M+1)(M+q)}$. The likelihood function for subject i is computed using the forward-backward algorithm [Rab89] as follows. For subject $i = 1, \ldots, n$, define the forward variables for $k = 1, \ldots, K-1$, and $m = 1, \ldots, M$,

$$\begin{aligned}\alpha_i^{T_k}(m; \boldsymbol{\theta}_i) \triangleq P_{\boldsymbol{\theta}_i}\Big\{ Y_i(T_1), \ldots, Y_i(T_k), S_i(T_k) = m \Big| \boldsymbol{X}_i(T_1) = \mathbf{x}_{T_1}, \\ \ldots, \boldsymbol{X}_i(T_k) = \mathbf{x}_{T_k} \Big\}.\end{aligned}$$

The initialization and recursion formulas are defined as

$$\begin{aligned}\alpha_i^{T_1}(m; \boldsymbol{\theta}_i) &= \pi_i(m; \boldsymbol{\theta}_i) g_{i,m}\{Y_i(T_1); \mathbf{x}_{T_1}, \boldsymbol{\theta}_i\}, \\ \alpha_i^{T_{k+1}}(m; \boldsymbol{\theta}_i) &= \left[\sum_{\ell=1}^{M} \alpha_i^{T_k}(\ell; \boldsymbol{\theta}_i) \{e^{(T_{k+1}-T_k)\mathbf{Q}_i}\}_{\ell,m}\right] g_{i,m}\{Y_i(T_{k+1}); \\ &\qquad \mathbf{x}_{T_{k+1}}, \boldsymbol{\theta}_i\},\end{aligned}$$

where $m = 1, \ldots, M$ and $k = 1, \ldots, K-1$. The negative log-likelihood for $\boldsymbol{\theta}_i$ is therefore $f_i(\boldsymbol{\theta}_i) = -log\left\{\sum_{m=1}^{M} \alpha_i^{T_K}(m; \boldsymbol{\theta}_i)\right\}$. Define the joint likelihood for $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$ to be $f(\boldsymbol{\theta}) = \sum_{i=1}^{n} f_i(\boldsymbol{\theta}_i)$.

To compute the conditional state probabilities in the HCTHMM, we need to generate a set of auxiliary backward variables analogous to the forward variables defined previously.

For subject $i = 1, \ldots, n$, define the backward variables for $k = 1, \ldots, K$, and $m = 1, \ldots, M$,

$$\begin{aligned}\beta_i^{T_k}(m;\boldsymbol{\theta}_i) \triangleq P_{\boldsymbol{\theta}_i}\Big\{ & Y_i(T_{k+1}), \ldots, Y_i(T_K) \Big| S_i(T_k) = m, \\ & \boldsymbol{X}_i(T_{k+1}) = \mathbf{x}_{T_{k+1}}, \ldots, \boldsymbol{X}_i(T_K) = \mathbf{x}_{T_K} \Big\}.\end{aligned}$$

The initialization and recursion formulas are

$$\begin{aligned}\beta_i^{T_K}(m;\boldsymbol{\theta}_i) =& 1, \\ \beta_i^{T_k}(m;\boldsymbol{\theta}_i) =& \sum_{\ell=1}^{M} \{e^{(T_{k+1}-T_k)\mathbf{Q}_i}\}_{m,\ell}\, g_{i,\ell}\{Y_i(T_{k+1}); \mathbf{x}_{T_{k+1}}, \boldsymbol{\theta}_i\} \\ & \beta_i^{T_{k+1}}(\ell;\boldsymbol{\theta}_i),\end{aligned}$$

where $m = 1, \ldots, M$. The probability of state m for subject i at time t is

$$\begin{aligned}\gamma_i^t(m;\boldsymbol{\theta}_i) &\triangleq P_{\boldsymbol{\theta}_i}\{S_i(t) = m | Y_i(T_1), \ldots, Y_i(T_K)\} \\ &= \frac{\alpha_i^t(m;\boldsymbol{\theta}_i)\beta_i^t(m;\boldsymbol{\theta}_i)}{\sum_{m=1}^{M} \alpha_i^t(m;\boldsymbol{\theta}_i)\beta_i^t(m;\boldsymbol{\theta}_i)},\end{aligned}$$

where $t = T_1, \ldots, T_k$, $m = 1, \ldots, M$, and $i = 1, \ldots, n$. The mean probability of state m among subjects in group j is thus

$$\begin{aligned}\phi_j(m;\boldsymbol{\theta}) &= \frac{1}{n_j} \sum_{\{i:G_i=j\}} \eta_i(m;\boldsymbol{\theta}_i),\ j = 1, \ldots, J, \\ \text{where } \eta_i(m;\boldsymbol{\theta}_i) &= \frac{1}{K}\sum_{k=1}^{K} \gamma_i^{T_k}(m;\boldsymbol{\theta}_i),\ m = 1, \ldots, M.\end{aligned}$$

The mean state probabilities $\phi_j(m;\boldsymbol{\theta})$ can be interpreted as the mean proportion of time spent in latent state m for subjects in group j, whereas $\eta_i(m;\boldsymbol{\theta}_i)$ represents the mean proportion of time spent in state m for subject i.

3.3.2 Consensus Optimization

If all sets of parameters are subject-specific, then the maximum likelihood estimates for the parameters can be obtained by minimizing $f(\boldsymbol{\theta})$ using the gradient-based methods which can be parallelized across subjects. However, in the general setting where parameters are shared across subgroups of subjects, such paralellization is no longer possible. Instead, we use the consensus optimization approach to obtain the maximum likelihood estimates in the HCTHMM, which is performed via the alternating direction method of multipliers (ADMM) [Boy11]. We use the Bayesian Information Criterion (BIC) to select the number of latent states M.

Let $\mathbf{D}$ denote a contrast matrix such that $\mathbf{D}\boldsymbol{\theta} = 0$ corresponds to equality of subgroup-specific parameters within each subgroup and equality of all population-level parameters across all subjects. The maximum likelihood estimator solves

$$\min_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) \ s.t. \ \mathbf{D}\boldsymbol{\theta} = 0.$$

For the purpose of illustration, suppose that: (i) the intercepts in the generalized linear models for state-dependent parameters are subject-specific; (ii) the initial state probabilities and the transition rate parameters are subgroup-specific; (iii) the slope parameters in the generalized linear models for state-dependent parameters are common across all subjects. Then $\mathbf{D}\boldsymbol{\theta} = 0$ is the same as restricting (i) $\mathbf{a}_{i_1} = \mathbf{a}_{i_2}$ for all $G_{i_1} = G_{i_2}$; (ii) $\mathbf{c}_{i_1} = \mathbf{c}_{i_2}$ for all $G_{i_1} = G_{i_2}$; (iii) $\mathbf{b}_{i,1} = \mathbf{b}_1$.

In our illustrative example, the maximum likelihood estimator solves

$$\begin{aligned} &\min_{\boldsymbol{\theta},\mathbf{z}} f(\boldsymbol{\theta}) \\ &s.t. \, \mathbf{A}_i \boldsymbol{\theta}_i = \mathbf{B}_i \mathbf{z}, \ i = 1, \dots, n, \end{aligned}$$

where $\mathbf{z}$ represents the set of all subgroup-specific and common parameters in $\boldsymbol{\theta}$ so the linear constraint $\mathbf{A}_i \boldsymbol{\theta}_i = \mathbf{B}_i \mathbf{z}$ is equivalent to $\mathbf{D}\boldsymbol{\theta} = 0$. The corresponding augmented Lagrangian is

$$\begin{aligned} L_\rho(\boldsymbol{\theta}, \mathbf{z}, \xi) &= f(\boldsymbol{\theta}) + \xi^T(\mathbf{A}\boldsymbol{\theta} - \mathbf{B}\mathbf{z}) + \frac{\rho}{2}\|\mathbf{A}\boldsymbol{\theta} - \mathbf{B}\mathbf{z}\|_2^2 \\ &= \sum_{i=1}^{n} \{f_i(\boldsymbol{\theta}_i) + \xi_i^T(\mathbf{A}_i\boldsymbol{\theta}_i - \mathbf{B}_i\mathbf{z}) + \frac{\rho}{2}\|\mathbf{A}_i\boldsymbol{\theta}_i - \mathbf{B}_i\mathbf{z}\|_2^2\}, \end{aligned}$$

where $\xi = [\xi_1', \dots, \xi_n']$,

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{A}_2 & \cdots & 0 \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & \mathbf{A}_n \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_n \end{bmatrix}.$$

Here ξ are the Lagrange multipliers and ρ is a pre-specified positive penalty parameter. Let

$\tilde{\boldsymbol{\theta}}_n^{(v)}, \tilde{\mathbf{z}}_n^{(v)}, \tilde{\xi}_n^{(v)}$ be the v^{th} iterates of $\boldsymbol{\theta}, \mathbf{z}, \xi$. Then, at the iteration $v+1$, the ADMM algorithm updates are

$$\begin{aligned}
&\boldsymbol{\theta}-\text{update: } \nabla f(\tilde{\boldsymbol{\theta}}_n^{(v+1)}) + \mathbf{A}^T \tilde{\xi}_n^{(v)} + \rho \mathbf{A}^T (\mathbf{A}\tilde{\boldsymbol{\theta}}_n^{(v+1)} - \mathbf{B}\tilde{\mathbf{z}}_n^{(v+1)}) = 0,\\
&\mathbf{z}-\text{update: } \mathbf{B}^T \tilde{\xi}_n^{(v)} + \rho \mathbf{B}^T (\mathbf{A}\tilde{\boldsymbol{\theta}}_n^{(v+1)} - \mathbf{B}\tilde{\mathbf{z}}_n^{(v)}) = 0,\\
&\xi-\text{update: } \tilde{\xi}_n^{(v+1)} - \tilde{\xi}_n^{(v)} - \rho (\mathbf{A}\tilde{\boldsymbol{\theta}}_n^{(v+1)} - \mathbf{B}\tilde{\mathbf{z}}_n^{(v+1)}) = 0,
\end{aligned}$$

where the most computationally expensive $\boldsymbol{\theta}$-update can be programmed in parallel across each $i = 1, \ldots, n$ as

$$\boldsymbol{\theta}_i - \text{update: } \nabla f_i(\tilde{\boldsymbol{\theta}}_{i,n}^{(v+1)}) + \mathbf{A}_i^T \tilde{\xi}_{i,n}^{(v)} + \rho \mathbf{A}_i^T (\mathbf{A}_i \tilde{\boldsymbol{\theta}}_i^{(v+1)} - \mathbf{B}_i \tilde{\mathbf{z}}_n^{(v)}) = 0.$$

The gradients $\nabla f_i(\cdot)$ for subject i's HMM parameters can be computed using Fisher's identity [Cap05] based on the efficient EM algorithm proposed in [Liu15]. The details are included in Appendix B. The use of gradients is needed in our model both due to the ADMM update and the covariate structure. Even when there are no covariates and all parameters are shared across subjects, the gradient method is still faster than the EM algorithm because M-step is expensive in the zero-inflated Poisson distribution.

3.3.3 Theoretical Properties

Define $\hat{\boldsymbol{\theta}}_n$ to be the maximum likelihood estimator for $\boldsymbol{\theta}$ and let $\boldsymbol{\theta}^\star$ denote its population-level analog. The following are the sufficient conditions to ensure: (i) almost sure convergence of $\hat{\boldsymbol{\theta}}_n$ to $\boldsymbol{\theta}^\star$ as $n \to \infty$, and (ii) numerical convergence of $\tilde{\boldsymbol{\theta}}_n^{(v)}$ to $\hat{\boldsymbol{\theta}}_n$ as $v \to \infty$.

(A0) The true parameter vector $\boldsymbol{\theta}^\star$ for the unconstrained optimization problem $\min_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$ is an interior point of Θ, where Θ is a compact subset of $\mathbb{R}^{\dim \boldsymbol{\theta}}$.

(A1) The constraint set $\mathscr{C} \triangleq \{\boldsymbol{\theta} \in \Theta; \mathbf{D}\boldsymbol{\theta} = 0\}$ is nonempty and for some $r \in \mathbb{R}$, the set $\{\boldsymbol{\theta} \in C; f(\boldsymbol{\theta}) \le r\}$ is nonempty and compact.

(A2) The observed time process $(T_k : k \in \mathbb{N})$ is independent of the generative hidden Markov process: the likelihood for the observed times do not share parameters with $\boldsymbol{\theta}$.

(A3) There exist positive real numbers $0 < \kappa^- \le \kappa^+ < 1$ such that for all subjects $i = 1, \ldots, n$, $\kappa^- \le P_i^{T_{k+1}-T_k}(m, \ell) \le \kappa^+$ for $k = 1, \ldots, K-1$ and $m, \ell = 1, \ldots, M$, almost surely and $g_{i,m}(y; \mathbf{x}) > 0$ for all $y \in$ supp Y for some $m = 1, \ldots, M$.

(A4) For each $\boldsymbol{\theta} \in \Theta$, the transition kernel indexed by $\boldsymbol{\theta}$ is Harris recurrent and aperiodic. The transition kernel is continuous in $\boldsymbol{\theta}$ in an open neighborhood of $\boldsymbol{\theta}^\star$.

(A5) The hidden Markov model is identifiable up to label switching of the latent states.

(A6) Denote by $e_{\min}(\boldsymbol{\theta})$ and $e_{\max}(\boldsymbol{\theta})$ to be the smallest and largest eigenvalues of $\nabla^2 f(\boldsymbol{\theta})$. There exist positive real numbers ϵ and $\varrho_- \leq \varrho_+$ such that $\inf_{\theta:\|\theta-\theta^\star\|_2<\epsilon} e_{\min}(\boldsymbol{\theta}) \geq \varrho_- > 0$ and $\sup_{\theta:\|\theta-\theta^\star\|_2<\epsilon} e_{\max}(\boldsymbol{\theta}) \leq \varrho_+ < \infty$.

Assumption (A0)-(A2) are mild regularity conditions whereas (A3) - (A5) are standard in hidden Markov models [Dou04]; together they ensures that the model is well-defined. Assumption (A4) avoids non-standard asymptotic behavior associated with non-smooth functionals. Assumption (A6) states that the smallest and largest eigenvalues for the Hessian of the negative log likelihood function are bounded away from 0 and infinity in an open neighborhood of the true parameter values, indicating the local Lipschitz differentiability and convexity, which are used in [Shi14] to establish the numerical convergence in the ADMM algorithm.

Theorem 3.3.1. *Under assumptions (A0) - (A5), as $T_i \to \infty$ for $i = 1, \ldots, n$,*

(i) $\hat{\boldsymbol{\theta}}_n$ converges to $\boldsymbol{\theta}^\star$ almost surely as $n \to \infty$,

(ii) suppose (A6) also holds, then $\tilde{\boldsymbol{\theta}}_n^{(v)}$ converges numerically to $\hat{\boldsymbol{\theta}}_n$ in an open neighborhood of $\boldsymbol{\theta}^\star$ as $v \to \infty$.

The first part of Theorem 1 states the almost sure convergence of the constrained maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$ to the true parameter value $\boldsymbol{\theta}^\star$. This can be shown using the uniform convergence results of the log likelihood [Dou04] for each subject-specific hidden Markov model, along with the feasibility assumption (A1) and identifiability assumption (A5). The second part of Theorem states the numerical convergence of the ADMM algorithm. This is anticipated by [Boy11] which identifies general conditions for the numerical convergence of the residual, the dual variable, and the objective function. [Shi14] extended the convergence to the primal variable by adding the Lipschitz continuity and convexity assumptions. The details for those assumptions, as well as the proof for Theorem 1, are included in Appendix B. For $i = 1, \ldots, n$, define the estimator for the mean proportion of time in state m in group k as $\hat{\phi}_{j,n}(m)$, $j = 1, \ldots, J, m = 1, \ldots, M$, and the estimator for the mean proportion of time in state m as $\hat{\eta}_{i,k_i}$ where k_i is the number of observed time points for subject i. The following result characterizes the limiting behavior of the estimated time in each state.

Theorem 3.3.2. *Under (A0) - (A5), as $k_i \to \infty$ for $i = 1, \ldots, n$, $n_j \to \infty$ for $j = 1, \ldots, J$,*

(i) $\hat{\phi}_{j,n}(m; \hat{\boldsymbol{\theta}}_n)$ *converges to* $\mu_j(m; \boldsymbol{\theta}^\star)$ *almost surely,*

(ii) $\frac{\hat{\phi}_{j,n}(m;\hat{\boldsymbol{\theta}}_n) - \mu_j(m;\hat{\boldsymbol{\theta}}_n)}{\sqrt{\sigma_j^2(m;\hat{\boldsymbol{\theta}}_n)/n_j}}$ *converges in distribution to a standard normal random variable, where* $\mu_j(m; \hat{\boldsymbol{\theta}}_n) = E[\hat{\eta}_{i,k_i}(m; \hat{\boldsymbol{\theta}}_n)]$, $\sigma_j^2(m; \hat{\boldsymbol{\theta}}_n) = Var[\hat{\eta}_{i,k_i}(m; \hat{\boldsymbol{\theta}}_n)]$ *for all i such that* $G_i = j$.

A proof of the preceding result is given in Appendix B, which follows from the almost sure convergence of a bounded continuous function and the central limit theorem. In principle, $\mu_j(m; \hat{\boldsymbol{\theta}}_n)$ can be obtained from the limiting distribution of a stationary continuous-time Markov chain, which is determined by the transition as a function of $\hat{\boldsymbol{\theta}}_n$. However, it is generally not easy to compute the standard error analytically for the estimated mean state probabilities. Instead, we use a stratified nonparametric bootstrap [Efr92] in which we resample subjects with replacement from each subgroup.

3.4 Simulation experiments

We study the finite sample performance of the proposed estimator for the state probabilities using a suite of simulation experiments. We simulate minute-by-minute activity counts of length $T \sim \text{Uniform}(500, 2500)$ for $n = 20$ and $n = 200$ subjects, where half of the subjects are male (Group 1) and the other half female (Group 2). The intervals between consecutive time points are independently drawn from $\{1, 2, \ldots, 10\}$ with equal probabilities. For each subject, we assume 2/5 of the observations are from weekends and 3/5 of the observations are from weekdays.

The activity counts are generated using a three state continuous-time zero-inflated Poisson hidden Markov model. We assume that during the weekend the log mean activity decreases by 10%, 20%, 30% in states 1, 2, 3 respectively, while the log odds of zero in state 1 increases by 10%, so that

$$\begin{aligned}
\log\left\{\frac{\delta_i(\mathbf{x})}{1-\delta_i(\mathbf{x})}\right\} &= b_{i,0,0} + 0.1 \times \mathbf{I}\{\text{Weekend}\}_i^t, \\
\log\left\{\lambda_{i,1}(\mathbf{x})\right\} &= b_{i,1,0} - 0.1 \times \mathbf{I}\{\text{Weekend}\}_i^t, \\
\log\left\{\lambda_{i,2}(\mathbf{x})\right\} &= b_{i,2,0} - 0.2 \times \mathbf{I}\{\text{Weekend}\}_i^t, \\
\log\left\{\lambda_{i,3}(\mathbf{x})\right\} &= b_{i,3,0} - 0.3 \times \mathbf{I}\{\text{Weekend}\}_i^t,
\end{aligned}$$

where $b_{i,0,0} \overset{iid}{\sim} N(-1, 0.1^2)$, $b_{i,1,0} \overset{iid}{\sim} N\left\{\log(50), 0.1^2\right\}$, $b_{i,2,0} \overset{iid}{\sim} N\left\{\log(300), 0.1^2\right\}$,

$b_{i,3,0} \overset{iid}{\sim} N\left\{\log(700), 0.1^2\right\}$ are subject-specific intercepts; the weekend effect is assumed to be common across all subjects.

The initial probabilities for male are $(U_1, U_2, 1-U_1-U_2)$, $U_1, U_2 \overset{iid}{\sim}$ Uniform$(0.2, 0.4)$; for female, the initial probabilities are $(U_3, U_4, 1-U_3-U_4)$, $U_3 \overset{iid}{\sim}$ Uniform$(0.6, 0.8)$, $U_4 \overset{iid}{\sim}$ Uniform$(0.1, 0.2)$. The transition rate matrix for male is

$$\begin{bmatrix} -U_5-U_6 & U_5 & U_6 \\ U_7 & -U_7-U_8 & U_8 \\ U_9 & U_{10} & -U_9-U_{10} \end{bmatrix},$$

where $U_5, \ldots, U_{10} \overset{iid}{\sim}$ Uniform$(0.05, 0.15)$; for female, the transition rate matrix is

$$\begin{bmatrix} -U_{11}-U_{12} & U_{11} & U_{12} \\ U_{13} & -U_{13}-U_{14} & U_{14} \\ U_{15} & U_{16} & -U_{15}-U_{16} \end{bmatrix},$$

where $U_{11}, U_{12} \overset{iid}{\sim}$ Uniform$(0.05, 0.1)$, $U_{13}, U_{15} \overset{iid}{\sim}$ Uniform$(0.3, 0.4)$, and $U_{14}, U_{16} \overset{iid}{\sim}$ Uniform$(0.1, 0.2)$.

Table 3.1 The bias and standard error for the estimated population and subgroup-specific parameters in HCTHMM. The metric is Euclidean norm for population parameters, and Frobenius norm for subgroup-specific parameter vectors.

Parameter	n=20	n=200
	Bias (s.e.)	Bias (s.e.)
Population parameters:		
Slope for State 1 zero odds	.0021 (.0422)	.0032 (.0134)
Slope for State 1 Poisson mean	.0001 (.0034)	.0001 (.0011)
Slope for State 2 Poisson mean	.0016 (.0053)	.0019 (.0018)
Slope for State 3 Poisson mean	.0036 (.0120)	.0046 (.0044)
Subgroup-specific parameters:		
Initial probabilities (Male)	.0169 (.0022)	.0165 (.0008)
Initial probabilities (Female)	.0152 (.0069)	.0284 (.0038)
Transition rates (Male)	.0038 (.0029)	.0011 (.0006)
Transition rates (Female)	.0099 (.0089)	.0016 (.0012)

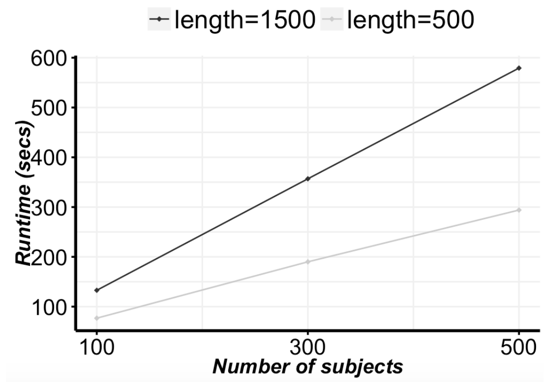

Figure 3.1 Average runtime (seconds) using eight cores on a Unix cluster for each ADMM iteration in different configurations (number of subjects, length of each series).

Table 3.2 Comparisons on the mean coverage probability for the 95% and 99% bootstrap confidence intervals for the mean proportion of time in each latent state between subject-specific HMM and HCTHMM (n = 200).

State	Subject-specific HMM		HCTHMM	
	Male	Female	Male	Female
95% C.I.:				
1	.942	.942	.942	.950
2	.944	.938	.950	.960
3	.946	.936	.960	.944
99% C.I. :				
1	.982	.988	.986	.994
2	.986	.984	.986	.988
3	.978	.980	.994	.992

Table 3.1 shows the bias and standard error of the estimators for different hierarchies of parameters in the HCTHMM via 500 simulations. In both cases, the biases are small due to the fact that the length of each individual series is large. As the sample size increases, the standard errors become smaller which is expected. Figure 3.1 shows the average runtime (seconds) for each ADMM iteration scales linearly with the number of subjects. It generally takes some 30 to 100 iterations for the algorithm to converge. Table 4.2 compares the mean coverage probability of the 95% and 99% bootstrap confidence intervals for the mean

proportion of time in each latent state between a baseline subject-specific HMM and the proposed HCTHMM when the sample size is 200. As we can see, the baseline subject-specific HMM suffers undercoverage (coverage probability smaller than nominal level), while the proposed HCTHMM recovers the nominal level on average in both the 95% and 99% cases.

3.5 Application

The motivating application is a human physical activity data set from the 2003 - 2004 National Health and Nutrition Examination Survey (NHANES), which is publicly available at the National Center for Disease Control (CDC) website `https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/PAXRAW_C.htm`. There are 7,176 participants in the study, and for each participant we have minute-by-minute activity counts for up to seven days. As the subjects were supposed to remove the accelerometer when washing, there are prolonged intervals during the day when accelerometer readings are zeros. We further impose the following two inclusion / exclusion criteria,

- Subjects whose age is between 20 and 60 are included.
- Subjects with very few measurements are excluded.

The first criterion specifies the scope of inference. The second criterion exclude subjects with very few non-missing data available ($<$ 500 minutes out of 7 days). There are 2,467 subjects who satisfy both conditions, which constitute more than 95% of those whose age is between 20 and 60. Further, we split those subjects by their baseline characteristics (gender, age) into 4 subgroups. Subgroup 1 consists of 608 male subjects with age from 20 to 40; subgroup 2 consists of 557 male subjects with age from 40 to 60; subgroup 3 consists of 712 female subjects with age from 20 to 40; and subgroup 4 consists of 590 female subjects with age from 40 to 60.

Table 3.3 The definition of missing interval in terms of consecutive minutes of zeros in the literature on human activity.

Literature	Definition of missing
[Cra04]	30 minutes
[Cat05]	20 minutes
[Tro08]	60 minutes
[Rob10]	20 minutes
[Eve11]	20 minutes
[Sch15]	60 minutes
[LG16]	20 minutes

Table 3.3 summarizes the related work on the length of an extended period of zero activity counts to be defined as missingness. In this paper, we choose to define missingness as a sustained interval of greater than or equal to 20 consecutive zero activity counts, which is the most commonly used criterion in the literature. Most missingness occurs between 10 pm to 8 am, which is the sleep time for most of the subjects. There is still sporadic missingness during other periods of time in the day, which may correspond to activities like swimming or bathing. The missingness periods are removed during the data preprocessing. The average proportion of zeros after removing the missingness is around 25%, so that zero-inflation is still an issue to be considered in the modeling. In the data preprocessing, activity counts greater than 1,500 ($< 5\%$) are truncated at 1,500 to ensure the numeric stability of the fitting algorithm.

Table 3.4 Summary of BIC from model selection. In type I models, all parameters are subject-specific. In type II models, all parameters are subject-specific except the slopes, which are population parameters. In type III models, the intercepts are subject-specific; the slopes are population; the initial probabilities and transitions are subgroup-specific. In type IV models, all parameters are subgroup-specific except the intercepts, which are subject-specific.

Model specifications	BIC
5 states, type I	248,009,082
6 states, type I	202,804,081
7 states, type I	203,261,150
6 states, type II	200,457,217
6 states, type III	198,949,672
6 states, type IV	198,807,080

To apply the HCTHMM model on the activity counts data, we need to select the number of latent states as well as the hierarchy for different sets of the parameters. The weekend effect is adjusted for in the Poisson and zero-inflated Poisson regression on the activity counts in each latent state. By the minimum BIC criterion as shown in Table 3.4, we select the type IV HCTHMM with 6 latent states, where the intercepts in the state-dependent generalized linear models for logit zero proportion in state 1 and log Poisson means in the all states are subject-specific, while the initial probabilities, transition rates, and the slopes in the state-dependent generalized linear models are subgroup-specific. This final model indicates the baseline zero proportion and mean activity counts in each latent state vary across subjects. For all the other parameters, the between-subgroup variability is more prominent than the within-subgroup variability. Figure 3.2 shows the 99% confidence interval for the estimated proportion of time spent in latent activity states for each subgroup in 03 - 04 NHANES. There are several interesting findings. First, younger men spend less time in the low intensity activity states (state 1, 2) than older men and women. Second, men spend less time than women in the medium intensity activity states (state 3, 4). Third, men spend more time than women in the high intensity activity states (state 5, 6). To validate the results, we apply the HCTHMM methodology to 05 - 06 NHANES, which has the same study setup and data structure as the 03 - 04 NHANES. Figure 3.3 shows the 99% confidence interval for the estimated proportion of time spent in latent activity states for each subgroup in 05 - 06 NHANES, which has a similar pattern as seen in 03 - 04 NHANES.

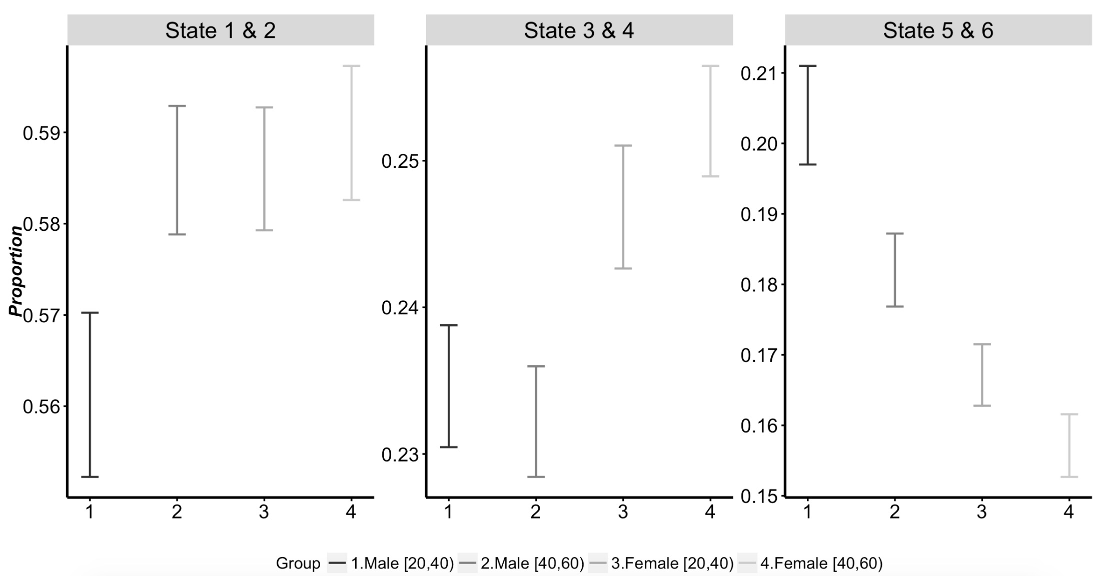

Figure 3.2 The 99% bootstrap confidence intervals for the estimated proportion of time spent in latent activity states by subgroup in 03 - 04 NHANES.

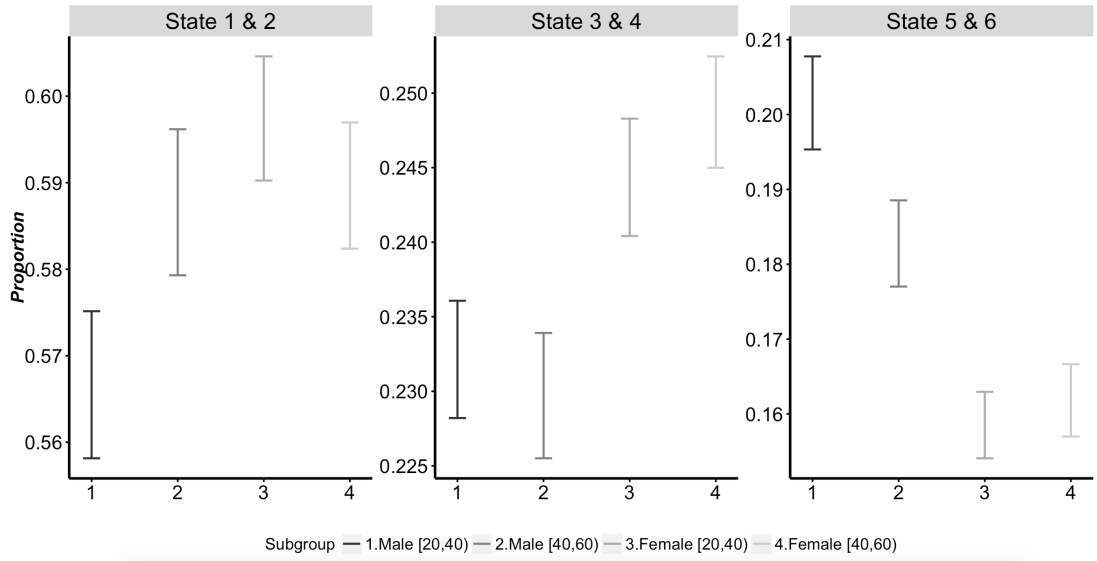

Figure 3.3 The 99% bootstrap confidence intervals for the estimated proportion of time spent in latent activity states for each subgroup in 05 - 06 NHANES.

3.6 Conclusions

We propose HCTHMM to be valid inference strategy for the longitudinal activity data. Within this framework, we can estimate the mean state probabilities for different subgroups of subjects as well as quantify the uncertainty. Our findings are consistent with previous literature on human physical activity [Met08; Tro08; Han12; Xia14], which indicated that the physical activity can be classified into different categories by intensity, and that the activity level decreases as a result of aging. Moreover, women tend to spend more time in lighter intensity activity, whereas younger men tend to have periods of higher intensity activities.

CHAPTER

4

LATENT STATE MODELS FOR PRECISION MEDICINE

4.1 Introduction

A treatment regime is a set of decision rules that determines a personalized treatment plan based on evolving treatment and covariate history [Mur03; CM14; Tsi19]. An optimal treatment regime maximizes the mean of some cumulative measure of patient health across a target population. Thus, there is keen interest in the development of statistical methodology for the estimation of optimal treatment regimes both to inform clinical practice and to generate new hypotheses about heterogeneous treatment effects [AI16; WA18]. Seminal methods for estimating optimal treatment regimes from observational or randomized studies included g-estimation [Rob97; Mur03; Rob04], Q-learning and its variants [Mur05; Moo07; Hen10; Sch14; Moo14a; Son15; Tay15; Zha18; Ert19], inverse probability weighting for the estimation of treatment effects of treatment regimes [Rob99; Mur01; Laa06; Rob08]. More recently, there has been a surge of research on extending these methods to make them more flexible, e.g., through the use of machine learning methods [Zha12b; Zha12a; RL12; Moo14a; Zha09; Zha15; LZ15; LL16; Xu16; WA18; Tao18; Jia19; Luc19; Liu19] or to allow them to work with high-dimensional feature spaces or other complex data structures [Lu13; MQ14; Tia14; Son15; Cia15; Cia16; LS17; Shi18; Ert19; Wal19; Shi19]. While the literature on treatment regimes is rich and growing rapidly, the type of data to which these methods applies is restrictive. Existing methods for finite-time-horizon decision problems require that one be able to align patient treatment decisions in time and that the conditional average treatment effect at each decision point can be estimated either using regression or weighting methods. For indefinite-time-horizon problems, existing methods for estimating optimal treatment regimes require that the data-generating distribution have sufficient structure to allow pooling of data over time points and extrapolation to future decisions, e.g., the data might be assumed to follow a contextual bandit or a homogeneous Markov decision process [TM17; Ert19; Luc19; Lia19]. Thus, existing methods do not apply to observational data when there are frequent and irregularly spaced treatment changes as patients cannot be properly aligned nor can the data be reasonably assumed to be Markov. For an example of such data, see Figure 4.1 which displays patient treatment histories for a subset of patients from the STEP-BD observational care pathway.

We propose a method for estimating an optimal treatment regime in the indefinite-time-horizon setting when data are irregularly spaced, contain multiple treatment changes, and cannot be assumed to be Markov. Motivated by the underlying clinical science of bipolar disorder and other episodic chronic illnesses, we assume that a patient's health status is dictated by a latent (unobserved) state and a subset of their observable data; we assume that conditional on current patient information and their latent state, the evolution of a patient's health status is Markov. Treatment is allowed to affect the transition dynamics of the latent process as well as patient observables. We show that under this model the optimal treatment regime is determined by the so-called information state which comprises the conditional distribution of the latent state and current patient measurements. We subsequently derive estimators of the optimal treatment regime and establish their asymptotic operating characteristics.

The proposed model is an example of a partially observable Markov decision process [POMDP Mon82]. POMDPs have been studied extensively in the computer science literature with applications in robotics, scheduling, videogames, and wildlife management [see, for example, Kae98; Cas98; JP03; Han98b; Ji07; SB18]. The primary contributions of this work include: a theory-driven construction of the latent-process model, the application of POMDPs to episodic chronic mental illness, and the development valid statistical inference for clinically relevant estimands in this context. The proposed methodology is extensible and could be ported for estimation and inference with optimal treatment regimes in other contexts having complex treatment and observation patterns, e.g., mobile-health.

The remainder of this manuscript is organized as follows. In Section 2, we formally introduce the latent state model and prove that the information state is sufficient for the optimal treatment regime. In Section 3, we review estimation of optimal treatment regimes under a Markov Decision Process model. In Section 4, we derive estimators of the optimal treatment regime based on a data-driven transformation of the observed process which makes it approximately Markov and thus amenable to the methods reviewed in Section 3. In Section 5, we derive the asymptotic distributions of the proposed estimators. In Section 6, we study the finite sample performance of the proposed methods through an extensive suite of simulation experiments. In Section 7, we provide an illustrative application using the observational care pathway of the STEP-BD bipolar disorder study. We provide a concluding discussion in Section 8.

4.2 Setup and preliminary results

We use uppercase letters, e.g., $\mathbf{X}$, T and A, to denote random variables, and lower case letters, e.g., $\mathbf{x}$, t and a, to denote instances of these random variables. The symbol '$\triangleq$' is used to distinguish definitions from equalities. The observed data are assumed to comprise n i.i.d. copies, one per patient, of the trajectory $\left\{[T^j, A(T^j), \mathbf{X}(T^j)]\right\}_{j=1}^{J}$, where $J \in \mathbb{Z}_+$ is the number of clinic visits, $0 = T^1 < T^2 < \cdots < T^J \leq 1$ encode clinic visit times, $A(T^j) \in \mathscr{A} = \{1, \ldots, L\}$ denotes the assigned treatment during period $[T^j, T^{j+1})$, and $\mathbf{X}(T^j) \in \mathbb{R}^p$ denotes a patient's health status at time T^j, $j = 1, \ldots, J$. Thus, both the number and timing of clinic visits are treated as random quantities. Let $\mathbf{H}^1 \triangleq \left\{T^1, \mathbf{X}(T^1)\right\}$ and $\mathbf{H}^j \triangleq \left\{\mathbf{H}^{j-1}, T^j, A(T^{j-1}), \mathbf{X}(T^j)\right\}$ so that $\mathbf{H}^j$ contains the patient history available to the decision maker at clinic visit T^j.

Let $\mathscr{D}$ denote the space of probability distributions over $\mathscr{A}$ (i.e., the L-dimensional probability simplex). We encode elements $\mathbf{d} \in \mathscr{D}$ vectors in $[0,1]^L$ so that d_a represents the probability of selecting action $a \in \mathscr{A}$ under $\mathbf{d}$. A treatment regime in this setting is a sequence of decision rules $\Pi = \left\{\pi^j\right\}_{j \geq 1}$, one per clinic visit, with $\pi^j : \operatorname{supp} \mathbf{H}^j \to \mathscr{D}$ so that under Π a patient presenting with $\mathbf{H}^j = \mathbf{h}_j$ at clinic visit j would receive treatment recommendation a with probability $\pi_a^j(\mathbf{h}_j)$. While the observed data comprise finite patient trajectories, we are interested in estimating treatment regimes that can be applied indefinitely; that is, they can be used to provide treatment recommendations for as long as the patient is receiving care. To this end, we consider treatment regimes composed of decision rules $\pi^j = \rho \circ f^j$, where $\mathbf{f} = \left\{f^j\right\}_{j \geq 1}$ are summary functions $f^j : \operatorname{supp} \mathbf{H}^j \to \mathscr{S} \subseteq \mathbb{R}^q$ so that $\mathbf{S}^j = f^j(\mathbf{H}^j)$ is a summary of patient history $\mathbf{H}^j$, and $\rho : \mathscr{S} \to \mathscr{D}$ is a stationary decision rule acting on patient summaries. Write $\boldsymbol{\pi} = \rho \circ \mathbf{f}$ to denote the composed regime $\pi^j = \rho \circ f^j$ for all $j \geq 1$. We shall show below that restricting attention to composed regimes of this type incurs no loss in generality. Furthermore, because ρ remains fixed in this representation, the regime can be vetted for clinical validity by domain experts when the summary functions provide 'qualitatively similar' summaries of patient history (we show that the natural choice of summary function in our domain produces such summaries).

The immediate utility associated with history $\mathbf{H}^j$ and treatment A^j is $U^j = \mathscr{U}(\mathbf{S}^j, A^j) \in \mathbb{R}$ and is assumed to depend on the history only through its summary. Note that this incurs no loss in generality as the summary function can always be chosen so that this assumption holds. Let $\mathbb{E}^{\pi}$ denote the probability distribution induced by following the treatment recommendations given by Π. We consider the following two measures of cumulative utility: (i) discounted mean utility

$$V_{\text{dis}}(\pi) \triangleq \mathbb{E}^{\pi} \sum_{j\geq 1} \left(\gamma^{j-1} U^j\right),$$

where $\gamma \in [0,1)$ is a discount factor; and
(ii) average utility

$$V_{\text{ave}}(\pi) \triangleq \lim_{J\to\infty} \mathbb{E}^{\pi} \left(\frac{1}{J} \sum_{j=1}^{J} U^j \right).$$

These two cumulative measures are used almost exclusively in indefinite decision problems [Pow07; Bus17; SB18] though the proposed methods could be extended to hyberbolic discounting or other notions of cumulative utility [Fed19]. We write $V(\pi)$ without a subscript to generically denote either of these cumulative utility measures. We say that π^{opt} is optimal with respect to V if $V(\pi^{\text{opt}}) \geq V(\pi)$ for all π. Without imposing additional structure on the data-generating model, it is not possible in general to identify π^{opt} from data collected over a finite time-horizon even as $n \to \infty$.

We assume that there exists a (latent) Markov process $M(t) \in \{1, \dots, K\}$ for all t which represents a critical component of a patient's health status, e.g., in bipolar disorder this might represent whether the patient is in a depressive, manic, hypo-manic, mixed, or stable episode. Furthermore, we assume that:

(A1) $\mathbf{X}^{j+1} \perp (\mathbf{H}^{j-1}, A^{j-1}) \big| \mathbf{X}^j, A^j, M(T^j)$,

so that the process is conditionally Markov given the latent state. The following the result shows that the conditional distribution of the latent state given the available history is sufficient for the optimal regime.

Lemma 4.2.1. *Assume (A1) and for $j \geq 1$ let $\mathbf{B}^j \in [0,1]^K$ be such that $B^j_\ell = P\left\{M(T^j) = \ell \big| \mathbf{H}^j\right\}$ for $\ell = 1, \dots, K$. Define $f^j(\mathbf{H}^j) \triangleq \left(\mathbf{B}^j, \mathbf{X}^j\right)$ and write $\mathbf{S}^j = f^j(\mathbf{H}^j)$ with $\mathscr{S} = \operatorname{supp} \mathbf{S}^j$. If U^j depends on $(\mathbf{H}^j, A^j)$ only through $(\mathbf{S}^j, A^j)$*[1] *then:*

[1] As noted previously, the utility should be a function of observables and thus $\mathbf{X}^j$ can always be defined so as to include U^j.

(i) $\left\{(\mathbf{S}^j, A^j, U^j)\right\}_{j\geq 1}$ *is a homogeneous Markov Decision Process;*

(ii) $\sup_{\rho:\mathscr{S}\to\mathscr{D}} V(\rho\circ\mathbf{f}) = \sup_{\Pi} V(\Pi)$.

Remark 4.2.1. *The summary* $\mathbf{S}^j$ *is minimally sufficient in that there exists generative models in which any further reduction of the history, e.g., learning a strategy which depends on* $\mathbf{R}^j = g^j(\mathbf{S}^j)$ *where* $||Cov(\mathbf{S}^j|\mathbf{R}^j)|| > 0$ $w.p.1$*, leads to degradation in the value of the learned strategy. See the Supplemental Material for a precise statement and example.*

Using the preceding result, we can characterize the optimal regime in terms of the Markov Decision Process (MDP) $\left\{(\mathbf{S}^j, A^j, U^j)\right\}_{j\geq 1}$. Because the states $\mathbf{S}^j$, $j \geq 1$ are not fully observed, we first construct estimators $\left\{\widehat{\mathbf{S}}_n^j\right\}_{j=1}^{J}$ of $\left\{\mathbf{S}^j\right\}_{j=1}^{J}$, and subsequently construct a plug-in estimator of the optimal regime.

4.3 Characterizing an optimal regime in an MDP

Recall that our approach is to transform the observed data so that (asymptotically) it is equivalent to a homogeneous Markov decision process (MDP). Thus, we briefly review two established methods for estimating an optimal treatment regime in an MDP. Our developments closely follow Murphy et al. [Mur16], Ertefaie [Ert19], and Luckett et al. [Luc19]. For a more general treatment of MDPs see Sutton & Barto [SB18] and Wiering & Van Otterlo [WVO12]. For the purpose of describing these methods, assume that the observed data are $\left\{(\mathbf{S}_i^j, A_i^j, U_i^j), j = 1, \ldots, J_i\right\}_{i=1}^{n}$, which consist of n independent trajectories from a homogeneous MDP. We assume that the states $\mathbf{S}^j \in \mathscr{S} \subset \mathbb{R}^q$ and, as in the previous section, there are a finite number of treatment options coded so that $A^j \in \{1, \ldots, L\}$ and the utilities $U^j \in \mathbb{R}$ are coded so that higher values are better. We present estimating equations for the optimal treatment regime with both the discounted and average reward criteria. Technical conditions needed for unbiasedness of these estimating equations and asymptotic normality of the resultant estimators applied to the transformed data, $\left\{(\widehat{\mathbf{S}}_{n,i}^j, A_i^j, U_i^j), j = 1, \ldots, J\right\}_{i=1}^{n}$, are provided in Section 5.

Define the discounted state-value function $\nu_{\text{dis}}(\pi, \mathbf{s}) = \mathbb{E}^{\pi}\left(\sum_{j\geq 0}\gamma^j U^{t+j}\middle|\mathbf{S}^t = \mathbf{s}\right)$ which does not depend on t because the MDP is assumed to be homogeneous. For any distribution R on $\mathscr{S}$, termed a reference distribution, define $V_{\text{dis}}^R(\pi) = \int \nu_{\text{dis}}(\pi, \mathbf{s})dR(\mathbf{s})$ then $V_{\text{dis}}(\pi) = V_{\text{dis}}^{R_0}(\pi)$, where R_0 is the initial state distribution. Because R_0 is unknown, one might take the empirical distribution of $\mathbf{S}^1$ or some other reference distribution constructed from historical data [see Luc19]. Define $\pi^{\text{opt},R} = \arg\max_{\pi\in\Pi} V_{\text{dis}}^R(\pi)$, where Π denotes a class of regimes of interest. In the discounted reward case it can be shown [e.g., Luc19] that the state-value function satisfies the following recursion

$$0 = \mathbb{E}\left[\frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j \middle| \mathbf{S}^j)}\left\{U^j + \gamma\, v_{\text{dis}}(\boldsymbol{\pi}, \mathbf{S}^{j+1}) - v_{\text{dis}}(\boldsymbol{\pi}, \mathbf{S}^j)\right\}\phi(\mathbf{S}^j)\right], \tag{4.1}$$

for all j and any $\phi : \mathscr{S} \to \mathbb{R}^d$, where the ratio $\pi_{A^j}(\mathbf{S}^j)/P(A^j|\mathbf{S}^j)$ is an importance sampling weight. Let $\mathscr{V} = \left\{ v(\mathbf{S}^j; \boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \Omega \subseteq \mathbb{R}^d \right\}$ be a parametric class of continuously differentiable maps from $\mathscr{S}$ into $\mathbb{R}^d$. For each $\boldsymbol{\pi} \in \Pi$, define $\boldsymbol{\alpha}^*(\boldsymbol{\pi})$ to be the solution to (4.1) at $\boldsymbol{\pi}$. An estimator $\widehat{\boldsymbol{\alpha}}_n(\pi)$ of $\boldsymbol{\alpha}^*(\pi)$ is given by the solution of the sample analogue of (4.1) with $\phi(\mathbf{S}^j) = \nabla_{\boldsymbol{\alpha}}\, v(\mathbf{S}^j; \boldsymbol{\alpha})$, i.e., the solution to

$$0 = \mathbb{P}_n \sum_{j=1}^{J-1}\left[\frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j \middle| \mathbf{S}^j)}\left\{U^j + \gamma\, v_{\text{dis}}(\mathbf{S}^{j+1}; \boldsymbol{\alpha}) - v_{\text{dis}}(\mathbf{S}^j; \boldsymbol{\alpha})\right\}\nabla_{\boldsymbol{\alpha}}\, v_{\text{dis}}(\mathbf{S}^j; \boldsymbol{\alpha})\right], \tag{4.2}$$

where $\mathbb{P}_n$ denotes the empirical measure. The estimated optimal regime is obtained by maximizing the estimated integrated state-value function over the class of regimes so that

$$\widehat{\pi}^R_{\text{dis},n} = \arg\max_{\pi \in \Pi} \int v_{\text{dis}}\left\{\mathbf{s}, \widehat{\alpha}_n(\pi)\right\} dR(\mathbf{s}).$$

Properties of this estimator—applied to data from a homogeneous MDP—are provided in Luckett et al. [Luc19]. We assumed for simplicity that $P(A^j|\mathbf{S}^j)$ was known, e.g., if the data were from a randomized clinical trial; if these were unknown, they could be estimated from the observed data, e.g., using a multinomial logistic regression [see also JL15; TB16; Han18, for related ideas and discussion].

An estimating equation for the average reward setting is derived using a similar strategy to the discounted case. For each π define the differential value

$$\delta(\boldsymbol{\pi}, \mathbf{s}) \triangleq \lim_{J \to \infty} \mathbb{E}^{\pi}\left[\sum_{j=1}^{J}\left\{U^j - V_{\text{ave}}(\pi)\right\}\middle| \mathbf{S}^1 = \mathbf{s}\right]$$

which is well-defined under regularity conditions provided in Section 5. Then it can be shown [[]e.g.,][]puterman2014markov,murphy2016batch,liao2019off that the $V_{\text{ave}}(\pi)$ satisfies the recursion

$$0 = \mathbb{E}\left[\frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j|\mathbf{S}^j)}\left\{U^j - V_{\text{ave}}(\pi) + \delta(\boldsymbol{\pi}, \mathbf{S}^{j+1}) - \delta(\boldsymbol{\pi}, \mathbf{S}^j)\right\}\psi(\mathbf{S}^j)\right], \tag{4.3}$$

for all j and any $\psi : \mathscr{S} \to \mathbb{R}^e$. Let $\mathscr{W} = \left\{\delta(\mathbf{s}; \beta) : \beta \in \mathscr{B} \subseteq \mathbb{R}^e\right\}$ be a class continuously differentiable maps from $\mathscr{S}$ into $\mathbb{R}^e$. An estimator $\widehat{V}_{\text{ave},n}(\pi)$ of $V_{\text{ave}}(\pi)$ is obtained by jointly solving the sample analog of (4.3) for β and $V_{\text{ave}}(\pi)$ with $\psi(\mathbf{s}) = (1, \nabla_\beta V(\mathbf{s};\beta)^\intercal)^\intercal$ so that $\widehat{V}_{\text{ave},n}(\pi)$, $\widehat{\beta}_n(\pi)$ solve

$$0 = \mathbb{P}_n \sum_{j=1}^{J-1} \left[\frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j|\mathbf{S}^j)} \left\{ U^j - V_{\text{ave}}(\pi) + \delta(\pi, \mathbf{S}^{j+1};\beta) - \delta(\pi, \mathbf{S}^j;\beta) \right\} \begin{pmatrix} 1 \\ \nabla_\beta \delta(\mathbf{S}^j;\beta) \end{pmatrix} \right]. \tag{4.4}$$

The estimated optimal regime is thus given by $\widehat{\pi}_{\text{ave},n} = \arg\max_{\pi \in \Pi} \widehat{V}_{\text{ave},n}(\pi)$.

Remark 4.3.1. *The remainder of this manuscript is focused on constructing the transformed process and examining the theoretical and empirical properties of the foregoing two estimators when applied to the transformed data. However, these are but two of many possible methods for estimating an optimal regimes with MDPs; these were chosen because they have been used previously in clinical applications and furthermore are simple, extensible, and amenable to statistical inference [[]for alternative approaches see][and references therein]szepesvari2010algorithms,powell2007approximate,sutton1998reinforcement.*

4.4 Estimating sufficient summary functions

Recall that the sufficient summary functions are given by $f_j = \left(\mathbf{B}^j, \mathbf{X}^j\right)$ for $j \geq 1$. As $\mathbf{X}^j$ is observed, constructing an estimator of f^j is tantamount to constructing an estimator of $\mathbf{B}^j$, the conditional distribution of the latent state given history $\mathbf{H}^j$. We develop an estimator of $\mathbf{B}^j$ under the assumption that the observables $\mathbf{X}^j$ evolve under a latent-state-dependent autoregressive process. This choice is motivated by the clinical theory underpinning bipolar disorder as well as its robustness and utility in modeling chronic illness [for additional discussion on time series and mechanistic models for bipolar disorder, see Dau09; Bon11; Moo12; Moo14b; Bon15; Hol16, and references therein].

We assume that the dynamics of the latent state $M(t)$ are described by the rate matrix $Q(a) \triangleq \left\{q_{k,\ell}(a)\right\}_{k,\ell=1,\ldots,K} \in \mathbb{R}^{K\times K}$ for each $a \in \{1,\ldots,L\}$, where

$$\begin{aligned} q_{k,k}(a) &\triangleq -\lim_{t\to 0^+} t^{-1} P\{M(T^{j+1}) \neq k | T^{j+1} - T^j = t, M(T^j) = k, A = a\}, \\ q_{k,\ell}(a) &\triangleq \lim_{t\to 0^+} t^{-1} P\{M(T^{j+1}) = \ell | T^{j+1} - T^j = t, M(T^j) = k, A = a\},\ k \neq \ell, \end{aligned}$$

from which it can be seen that $q_{k,k}(a) = -\sum_{\ell\neq k} q_{k,\ell}(a)$ for $k = 1,\ldots,K$ [see Liu15]. The transition rate matrix, also known as the infinitessimal generator [e.g., Pyk61a; Pyk61b;

Alb62], induces the following transition probabilities

$$P\{M(t')=\ell|M(t)=k,A=a\}=[\exp\{(t'-t)\cdot Q(a)\}]_{k,\ell},$$

for $k,\ell=1,\ldots,K$. To identify the transition rate matrix, we need to link the observed data to the latent state process. This link is established by the following assumption:

(A2) $\mathbf{X}^j \perp M(T^1),A^1,\mathbf{X}^1,\ldots,M(T^{j-1}),A^{j-1}\Big|M(T^j),\mathbf{X}^{j-1}$ for all $j\geq 1$.

Furthermore, we posit parametric models for the dynamics of the observed data and that these models have densities with respect to some dominating measure of the following form: the density of $\mathbf{X}^1$ given $M(T^1)=m^1$ is $p_{\mathbf{X}^1|M(T^1)}(\mathbf{x}^1|m^1;\Theta)$ which is indexed by $\Theta\in$ ˆ and the density of $\mathbf{X}^j$ given $M(T^j)=m^j$ and $\mathbf{X}^{j-1}=\mathbf{x}^j$ is $p_{\mathbf{X}^j|M(T^j),\mathbf{X}^{j-1}}(\mathbf{x}^j|m^j,\mathbf{x}^{j-1};\Gamma)$ which is indexed by $\boldsymbol{\gamma}\in\Gamma$. For example, a Gaussian autoregressive model with linear mean models would take the form:

$$p_{\mathbf{X}^1|M(T^1)}(\mathbf{x}^1|m^1;\Theta) \quad\propto\quad |\Sigma_{m^1}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{x}^1-\boldsymbol{\mu}_{m^1})^\intercal\Sigma_{m^1}^{-1}(\mathbf{x}^1-\boldsymbol{\mu}_{m^1})\right\},$$

where $\boldsymbol{\theta}=\left\{(\boldsymbol{\mu}_m,\Sigma_m)\right\}_{m=1,\ldots,K}$ are unknown parameters, and

$$p_{\mathbf{X}^j|M(T^j),\mathbf{X}^{j-1}}(\mathbf{x}^j|m^j,\mathbf{x}^{j-1};\Gamma)\propto$$
$$|\Sigma_{m^j}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{x}^j-\Psi_{m^{j-1}}\mathbf{x}^{j-1})^\intercal\Sigma_{m^{j-1}}^{-1}(\mathbf{x}^j-\Psi_{m^{j-1}}\mathbf{x}^{j-1})\right\},$$

where $\Gamma=\{(\Psi_m,\Sigma_m)\}_{m=1,\ldots,K}$. We use this model in our simulation experiments and application to the data from the STEP-BD trial.

Let $\boldsymbol{\varrho}=(\boldsymbol{\theta},\boldsymbol{\gamma})\in$ ˆ × ` denote the unknown parameters indexing the latent Markov process. It can be seen that $\mathbf{B}^j$ is determined by $\mathbf{H}^j$ and $\boldsymbol{\varrho}$, i.e., $\mathbf{B}^j=\mathbf{b}^j(\mathbf{H}^j,\varrho)$ where $\mathbf{b}^j$ is a deterministic map from $\mathbf{dom}\ \mathbf{H}^j\times(\Theta\times\Gamma)$ into the L-dimensional probability simplex. We construct an estimator $\widehat{\boldsymbol{\varrho}}_n$ via maximum likelihood implemented using the forward-backward algorithm [for a review see Rab89] and subsequently compute the plug-in estimator $\widehat{\mathbf{B}}_n^j=\mathbf{b}^j(\mathbf{H}^j,\widehat{\boldsymbol{\varrho}}_n)$ so that $\widehat{\mathbf{S}}_n^j=(\widehat{\mathbf{B}}_n^j,\mathbf{X}^j)$. The preceding estimator is used to convert i.i.d. trajectories of the form $\left\{[T_i^j,A_i(T_i^j),\mathbf{X}_i(T^j)]\right\}_{j=1}^{J_i}$ for $i=1,\ldots,n$ to trajectories drawn from an (approximate) homogeneous MDP $\left\{(\widehat{\mathbf{S}}_{i,n}^j,A_i^j,U_i^j)\right\}_{j=1}^{J_i}$ for $i=1,\ldots,n$ which can then be used with the estimators of an an optimal regime described in the previous section.

4.5 Theoretical properties

4.5.1 Consistency of estimated state probabilities

Consistency of the estimated latent state distribution is central to characterizing the large sample behavior of estimators of the optimal treatment regime constructed by solving the MDP estimating equations of Section 3. Consistency follows from existing results on maximum likelihood for latent Markov models and the continuous mapping theorem. We make the following assumptions.

(B1) Both the time process $(T^j : j \in \mathbb{N})$ and the number of time points J are independent of the latent process $\{M(t) : t \geq 0\}$.

(B2) The true parameter vector $\boldsymbol{\varrho}^*$ is an interior point of $\Theta \times \Gamma$, where $\Theta \times \Gamma$ is a compact subset of $\mathbb{R}^{\dim \boldsymbol{\varrho}}$.

(B3) For all $m^1 \in \{1, \ldots, K\}$, $P\{M(T^1) = m^1\} > 0$.

(B4) There exist non-trivial measures υ_1, υ_2 on $\mathscr{X}$ with

$$p_{\mathbf{X}^1|M(T^1)}(\cdot|m^1) \geq \upsilon_1(\cdot) \text{ for all } m^1 \in \{1, \ldots, K\},$$
$$p_{\mathbf{X}^j|M(T^j),\mathbf{X}^{j-1}}(\cdot|m^j, \mathbf{x}^{j-1}) \geq \upsilon_2(\cdot) \text{ for all } m^j \in \{1, \ldots, K\}, \mathbf{x}^j \in \mathscr{X}, j > 1,$$

where $\mathscr{X} = \operatorname{dom} \mathbf{X}^j$, $j \geq 1$.

(B5) For each $\boldsymbol{\varrho} \in \Theta \times \Gamma$, the transition kernel indexed by $\boldsymbol{\varrho}$ is stationary, Harris recurrent, and aperiodic [see AL06; MT12, for additional discussion of this assumption and its implications].

(B6) The transition kernel is continuous in $\boldsymbol{\varrho}$ in an open neighborhood of $\boldsymbol{\varrho}^*$.

(B7) The latent Markov process is identifiable up to label switching of the latent states [see All09; Gas13, for discussions of label-switching].

(B8) The log likelihood is twice continuously differentiable in $\boldsymbol{\varrho}$ and the Fisher information $\mathbf{I}(\boldsymbol{\varrho})$ is positive definite in an open neighborhood of $\boldsymbol{\varrho}^*$ [see Bic98; JP99; Dou04, for equivalent assumptions].

(B9) For any $\boldsymbol{\varrho}_1, \boldsymbol{\varrho}_2 \in \Theta \times \Gamma$, $||\mathbf{b}^j(\mathbf{h}^j, \boldsymbol{\varrho}_1) - \mathbf{b}^j(\mathbf{h}^j, \boldsymbol{\varrho}_2)|| \leq g^j(\mathbf{h}^j)||\boldsymbol{\varrho}_1 - \boldsymbol{\varrho}_2||$, for some integrable function $g^j : \operatorname{dom} \mathbf{H}^j \to \mathbb{R}$, $j \geq 1$.

The preceding assumptions are relatively mild and standard in hidden Markov models. Assumption (B1) ensures that the distribution of the visit times factors out of the likelihood for $\boldsymbol{\varrho}$, i.e., the time process and latent process do not share parameters. Assumptions (B2)-(B8) ensure that the model is well-defined and that the maximum likelihood estimators are regular [Ler92; Bic98; JP99; LGM00; DM01; Dou04]. Consistency and asymptotic normality of the maximum likelihood estimators in general autoregressive hidden Markov models have been established under the preceding conditions [Dou04]. Moreover, we will show that Assumption (B9) holds for the Gaussian autoregressive hidden Markov model in the Supplementary Material, which ensures the class is Donsker and thus consistency of the estimated state probabilities follows immediately.

Lemma 4.5.1. *Assume (A1) - (A2), (B1) - (B8), as $n \to \infty$:*

$$\begin{aligned}
&\sqrt{n}(\widehat{\boldsymbol{\varrho}}_n - \boldsymbol{\varrho}) \rightsquigarrow \mathcal{N}\{\mathbf{0}, \mathbf{I}(\boldsymbol{\varrho}^*)^{-1}\},\\
&\sqrt{n}\{\mathbf{b}^j(\mathbf{h}^j;\widehat{\boldsymbol{\varrho}}_n) - \mathbf{b}^j(\mathbf{h}^j;\boldsymbol{\varrho}^*)\} \rightsquigarrow \mathcal{N}\{\mathbf{0}, \nabla_{\boldsymbol{\varrho}}\mathbf{b}^j(\mathbf{h}^j;\boldsymbol{\varrho}^*)\mathbf{I}(\boldsymbol{\varrho}^*)^{-1}\nabla_{\boldsymbol{\varrho}}\mathbf{b}^j(\mathbf{h}^j;\boldsymbol{\varrho}^*)^{\intercal}\},
\end{aligned}$$

for each $\mathbf{h}^j \in \operatorname{dom}\mathbf{H}^j$. Furthermore, if (B9) holds, then for each fixed $j = 1, \ldots, J$, as $n \to \infty$,

$$\sup_{\mathbf{h}^j \in \operatorname{dom}\mathbf{H}^j} \left|\mathbf{b}^j(\mathbf{h}^j, \widehat{\boldsymbol{\varrho}}_n) - \mathbf{b}^j(\mathbf{h}^j, \boldsymbol{\varrho}^*)\right| \xrightarrow{p} 0.$$

4.5.2 Asymptotic properties in the discounted utility setting

We consider linear working models for the state-value function $v(\mathbf{s};\boldsymbol{\alpha}) = \phi(\mathbf{s})^{\intercal}\boldsymbol{\alpha}$, where ϕ is a finite-dimensional set of basis functions; these basis functions might comprise custom features informed by domain expertise as well as nonlinear expansions such as b-splines or radial basis functions. Using this functional form, the population-level estimating equation for the state value-function, i.e., (4.1) from Section 4.3, is given by

$$\Lambda_{\text{dis}}(\boldsymbol{\pi}, \boldsymbol{\alpha}) \triangleq \mathbb{E}\sum_{j=1}^{J-1}\left[\frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j \middle| \mathbf{S}^j)}\left\{U^j + \gamma\phi(\mathbf{S}^{j+1})^{\intercal}\boldsymbol{\alpha} - \phi(\mathbf{S}^j)^{\intercal}\boldsymbol{\alpha}\right\}\phi(\mathbf{S}^j)\right];$$

let $\boldsymbol{\alpha}^*(\boldsymbol{\pi})$ denote the solution to $\Lambda_{\text{dis}}(\boldsymbol{\pi}, \boldsymbol{\alpha}) = 0$. The sample analog using the estimated states is thus

$$\widehat{\Lambda}_{\text{dis},n}(\boldsymbol{\pi}, \boldsymbol{\alpha}) \triangleq \mathbb{P}_n\sum_{j=1}^{J-1}\left[\frac{\pi_{A^j}\left(\widehat{\mathbf{S}}_n^j\right)}{P\left(A^j \middle| \widehat{\mathbf{S}}_n^j\right)}\left\{\widehat{U}_n^j + \gamma\phi\left(\widehat{\mathbf{S}}_n^{j+1}\right)^{\intercal}\boldsymbol{\alpha} - \phi\left(\widehat{\mathbf{S}}_n^j\right)^{\intercal}\boldsymbol{\alpha}\right\}\phi\left(\widehat{\mathbf{S}}_n^j\right)\right],$$

where $\widehat{U}_n^j = \mathscr{U}(\widehat{\mathbf{S}}_n^j, A^j)$; let $\widehat{\boldsymbol{\alpha}}_n(\pi)$ denote a solution to $\widehat{\Lambda}_{\text{dis},n}(\pi, \boldsymbol{\alpha}) = 0$. For linear estimators, such a root always exists, however, below we require the weaker condition that $\widehat{\boldsymbol{\alpha}}_n(\pi)$ is an approximate root. Let $||\cdot||_F$ denote the Frobenius norm. We make the following assumptions.

(C1) For each $\pi \in \Pi$, $\boldsymbol{\alpha}^*(\pi)$ solves $\Lambda_{\text{dis}}(\pi, \boldsymbol{\alpha}) = 0$, where $\boldsymbol{\alpha}^*(\pi)$ is an interior point of Ω and Ω is a compact subset of $\mathbb{R}^{\dim \boldsymbol{\alpha}}$.

(C2) For each $\pi \in \Pi$, there exists a sequence of $\widehat{\boldsymbol{\alpha}}_n(\pi) \in \Omega$ such that $\widehat{\Lambda}_{\text{dis},n}\{\pi, \widehat{\boldsymbol{\alpha}}_n(\pi)\} = o_p(n^{-1/2})$.

(C3) Define $V_{\text{dis}}^R(\pi) \triangleq \int \phi(\mathbf{s})^\intercal \boldsymbol{\alpha}^*(\pi) dR(\mathbf{s})$, which attains its supremum at $\pi_{\text{dis}}^* \in \boldsymbol{\Pi}$.

(C4) There exists a sequence of $\widehat{\pi}_{\text{dis},n} \in \Pi$ such that $\widehat{V}_{\text{dis},n}^R(\widehat{\pi}_{\text{dis},n}) \geq \sup_{\pi \in \Pi} \widehat{V}_{\text{dis},n}^R(\pi) - o_p(1)$.

(C5) There exists a constant $c > 0$, such that

$$\boldsymbol{\omega}^\intercal \mathbb{E}\left[\frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j|\mathbf{S}^j)}\phi(\mathbf{S}^j)\left\{\phi(\mathbf{S}^j - \gamma\phi(\mathbf{S}^{j+1})\right\}^\intercal\right]\boldsymbol{\omega} \geq c\|\boldsymbol{\omega}\|_2^2,$$

for all $j \geq 1$ and $\boldsymbol{\omega} \neq \mathbf{0}$.

(C6) $\phi : \mathscr{S} \to \mathbb{R}^d$ is uniformly continuous, where $\mathscr{S} = \operatorname{dom} \mathbf{S}^j$, $j = 1, \ldots, J$, is compact, and J is finite almost surely. Furthermore, $\mathbb{E}||\mathbf{S}^j||^2 < \kappa$ for some $\kappa > 0$ and all $j \geq 1$.

(C7) For each $a \in \mathscr{A}$, $\mathbf{s} \in \mathscr{S}$, $j \geq 1$, $P\left(A^j = a|\mathbf{S}^j = \mathbf{s}\right) \geq \epsilon$, for some $\epsilon > 0$.

(C8) For each $j = 1, \ldots, J$, $\pi \in \Pi$, define

$$\mathbf{G}_{\text{dis}}^j\left(\pi, \mathbf{h}^j, a^j; \boldsymbol{\varrho}\right) \triangleq \frac{\pi_{a^j}\left\{b^j(\mathbf{h}^j; \boldsymbol{\varrho})\right\}}{P\left\{a^j|b^j(\mathbf{h}^j; \boldsymbol{\varrho})\right\}}\mathscr{U}\left\{b^j\left(\mathbf{h}^j; \boldsymbol{\varrho}\right), a^j\right\}\phi\left\{b^j\left(\mathbf{h}^j; \boldsymbol{\varrho}\right)\right\},$$

there exists a linear operator $\mathbf{W}_{\text{dis}}^j(\pi, \mathbf{h}^j, a^j; \boldsymbol{\varrho})$ such that $\mathbb{E}||\mathbf{W}_{\text{dis}}^j(\pi, \mathbf{H}^j, A^j; \boldsymbol{\varrho})||_F < \infty$ and, for all $\mathbf{h}^j \in \operatorname{dom} \mathbf{H}^j$ and $a^j \in \mathscr{A}$, the following expansion holds

$$\mathbf{G}_{\text{dis}}^j\left(\pi, \mathbf{h}^j, a^j; \widehat{\boldsymbol{\varrho}}_n\right) - \mathbf{G}_{\text{dis}}^j\left(\pi, \mathbf{h}^j, a^j; \boldsymbol{\varrho}^*\right) = \mathbf{W}_{\text{dis}}^j(\pi, \mathbf{h}^j, a^j; \boldsymbol{\varrho}^*)(\widehat{\boldsymbol{\varrho}}_n - \boldsymbol{\varrho}^*) + o_p(n^{-1/2}).$$

These conditions are standard for Z-estimators [VDVW96; Kos08]. Conditions (C1), (C2), (C6), and (C7) are used to establish the consistency of $\widehat{\boldsymbol{\alpha}}_n(\pi)$, while the addition of (C5) and (C8) are used to establish asymptotic normality. A sufficient condition for (C8) is that $\mathbf{G}_{\text{dis}}^j(\pi, \mathbf{h}^j, a^j; \boldsymbol{\varrho})$ is almost everywhere differentiable in $\boldsymbol{\varrho}$ in which case $\mathbf{W}_{\text{dis}}^j(\pi, \mathbf{h}^j, a^j; \boldsymbol{\varrho})$ can be chosen to be the gradient operator. We use (C3) and (C4) to show $\widehat{V}_{\text{dis},n}(\widehat{\pi}_{\text{dis},n}) \xrightarrow{p} V_{\text{dis}}(\pi_{\text{dis}}^*)$, which is a weaker but more general result than $\widehat{\pi}_{\text{dis},n} \xrightarrow{p} \pi^*_{\text{dis}}$. Convergence of $\widehat{\pi}_{\text{dis},n}$ generally requires that π^*_{dis} is a unique and well-separated maximizer of $V_{\text{dis}}(\pi)$, which need not hold for some commonly used classes of regimes [see Zha18].

Theorem 4.5.1. *Assume (A1) - (A2), (C1) - (C7), and that* Π *is finite. Then as* $n \to \infty$*:*

1. *for any fixed regime* π, $\widehat{\boldsymbol{\alpha}}_n(\pi) \xrightarrow{p} \boldsymbol{\alpha}^*(\pi)$*;*
2. $\widehat{V}_{\text{dis},n}(\widehat{\pi}_{\text{dis},n}) \xrightarrow{p} V_{\text{dis}}(\pi^*_{\text{dis}})$.

To define the limiting distribution of the estimated optimal value we make use of the following quantities:

$$
\begin{aligned}
C_1(\pi) &\triangleq \mathbb{E}\left[\sum_{j=1}^{J-1} \frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j|\mathbf{S}^j)} \phi(\mathbf{S}^j)\left\{\phi(\mathbf{S}^j) - \gamma\phi(\mathbf{S}^{j+1})\right\}^{\intercal}\right],\\
\widehat{C}_{1,n}(\pi) &\triangleq \mathbb{P}_n\left[\sum_{j=1}^{J-1} \frac{\pi_{A^j}(\widehat{\mathbf{S}}^j_n)}{P(A^j|\widehat{\mathbf{S}}^j_n)} \phi(\widehat{\mathbf{S}}^j_n)\left\{\phi(\widehat{\mathbf{S}}^j_n) - \gamma\phi(\widehat{\mathbf{S}}^{j+1}_n)\right\}^{\intercal}\right],\\
C_2(\pi,\widetilde{\pi}) &\triangleq \mathbb{E}\left[\sum_{j=1}^{J-1} \frac{\pi_{A^j}(\mathbf{S}^j)\widetilde{\pi}_{A^j}(\mathbf{S}^j)}{P^2(A^j|\mathbf{S}^j)}\left\{U^j + \gamma\phi(\mathbf{S}^{j+1})^{\intercal}\boldsymbol{\alpha}^*(\pi) - \phi\{\mathbf{S}^j\}^{\intercal}\boldsymbol{\alpha}^*(\pi)\right\}\right.\\
&\qquad\left.\left\{U^j + \gamma\phi(\mathbf{S}^{j+1})^{\intercal}\boldsymbol{\alpha}^*(\widetilde{\pi}) - \phi(\mathbf{S}^j)^{\intercal}\boldsymbol{\alpha}^*(\widetilde{\pi})\right\}\phi(\mathbf{S}^j)\phi(\mathbf{S}^j)^{\intercal}\right],\\
\widehat{C}_{2,n}(\pi,\widetilde{\pi}) &\triangleq \mathbb{P}_n\left[\sum_{j=1}^{J-1} \frac{\pi_{A^j}(\widehat{\mathbf{S}}^j_n)\widetilde{\pi}_{A^j}(\widehat{\mathbf{S}}^j_n)}{P^2(A^j|\widehat{\mathbf{S}}^j_n)}\left\{\widehat{U}^j_n + \gamma\phi(\widehat{\mathbf{S}}^{j+1}_n)^{\intercal}\widehat{\boldsymbol{\alpha}}_n(\pi) - \phi(\widehat{\mathbf{S}}^j_n)^{\intercal}\widehat{\boldsymbol{\alpha}}_n(\pi)\right\}\right.\\
&\qquad\left.\left\{\widehat{U}^j_n + \gamma\phi\left(\widehat{\mathbf{S}}^{j+1}_n\right)^{\intercal}\widehat{\boldsymbol{\alpha}}_n(\widetilde{\pi}) - \phi(\widehat{\mathbf{S}}^j_n)^{\intercal}\widehat{\boldsymbol{\alpha}}_n(\widetilde{\pi})\right\}\phi(\widehat{\mathbf{S}}^j_n)\phi(\widehat{\mathbf{S}}^j_n)^{\intercal}\right],\\
C_3(\pi) &\triangleq \mathbb{E}\left[\sum_{j=1}^{J-1} \mathbf{W}^j_{\text{dis}}(\pi,\mathbf{H}^j,A^j;\boldsymbol{\varrho}^*)\right],\\
\widehat{C}_{3,n}(\pi) &\triangleq \mathbb{P}_n\left[\sum_{j=1}^{J-1} \mathbf{W}^j_{\text{dis}}(\pi,\mathbf{H}^j,A^j;\widehat{\boldsymbol{\varrho}}_n)\right].
\end{aligned}
$$

Theorem 4.5.2. *Assume (A1) - (A2), (C1) - (C8), and that" is finite. The following results hold as* $n \to \infty$ *:*

1. $\sqrt{n}\{\widehat{V}^R_{\text{dis},n}(\pi) - V^R_{\text{dis}}(\pi)\} \rightsquigarrow \mathbb{B}(\pi)$, *where* $\mathbb{B}(\pi)$ *is a mean zero Gaussian process indexed by*

π with covariance

$$\mathbb{E}\{\mathbb{B}(\pi)\mathbb{B}(\widetilde{\pi})\} = \left[\int \phi(\mathbf{s})dR(\mathbf{s})\right]^{\intercal} C_1^{-1}(\pi)[C_2(\pi,\widetilde{\pi}) + C_3(\pi)\mathbf{I}^{-1}(\boldsymbol{\varrho}^*)C_3^{\intercal}(\widetilde{\pi})]$$
$$C_1^{-\intercal}(\widetilde{\pi})\left[\int \phi(\mathbf{s})dR(\mathbf{s})\right];$$

2. $\sqrt{n}\widehat{\sigma}_{\mathrm{dis},n}^{-1/2}(\pi)\left\{\widehat{V}_{\mathrm{dis},n}^{R}(\pi) - V_{\mathrm{dis}}^{R}(\pi)\right\} \rightsquigarrow \mathcal{N}(0,1)$, *where*

$$\widehat{\sigma}_{\mathrm{dis},n}^{-1/2}(\pi) = \left[\mathbb{P}_n\{\phi(\mathbf{s})\}^{\intercal}\widehat{C}_{1,n}^{-1}(\pi)[\widehat{C}_{2,n}(\pi,\pi) + \widehat{C}_{3,n}(\pi)\widehat{\mathbf{I}}_n^{-1}(\widehat{\boldsymbol{\varrho}}_n)\widehat{C}_{3,n}^{\intercal}(\pi)]\widehat{C}_{1,n}^{-\intercal}(\pi)\mathbb{P}_n\{\phi(\mathbf{s})\}\right]^{1/2}.$$

4.5.3 Asymptotic properties in the average utility setting

We derive the limiting distribution of the value function under a linear working model for the differential value $\delta(\mathbf{s},\boldsymbol{\beta}) = \phi(\mathbf{s})^{\intercal}\boldsymbol{\beta}$, where ϕ is a vector of features constructed from $\mathbf{s}$ as in the preceding section. Further, define $\boldsymbol{\zeta}(\pi) \triangleq \left\{V_{\mathrm{ave}}(\pi), \boldsymbol{\beta}\right\}$, $\psi_1(\mathbf{s}) \triangleq \left\{1, \phi(\mathbf{s})\right\}^{\intercal}$, and $\psi_2(\mathbf{s},\widetilde{\mathbf{s}}) \triangleq \left\{1, \phi(\mathbf{s}), \phi(\widetilde{\mathbf{s}})\right\}^{\intercal}$. The population estimating equation for the average utility, i.e., equation (4.3) in Section 4.3, under the posited model is

$$\Lambda_{\mathrm{ave}}\left\{\pi, \boldsymbol{\zeta}(\Pi)\right\} \triangleq \mathbb{E}\sum_{j=1}^{J-1}\left[\frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j|\mathbf{S}^j)}\left\{U^j - \psi_2(\mathbf{S}^j,\mathbf{S}^{j+1})^{\intercal}\boldsymbol{\zeta}(\pi)\right\}\psi_1(\mathbf{S}^j)\right];$$

define $\boldsymbol{\zeta}^*(\pi)$ as the solution to $\Lambda_{\mathrm{ave}}\left\{\pi, \boldsymbol{\zeta}(\Pi)\right\} = 0$. The sample analog is

$$\widehat{\Lambda}_{\mathrm{ave},n}\left\{\pi, \boldsymbol{\zeta}(\pi)\right\} \triangleq \mathbb{P}_n\sum_{j=1}^{J-1}\left[\frac{\pi_{A^j}(\widehat{\mathbf{S}}_n^j)}{P(A^j|\widehat{\mathbf{S}}_n^j)}\{\widehat{U}_n^j - \psi_2(\widehat{\mathbf{S}}_n^j,\widehat{\mathbf{S}}_n^{j+1})^{\intercal}\boldsymbol{\zeta}(\pi)\}\psi_1(\widehat{\mathbf{S}}_n^j)\right],$$

where $\widehat{U}_n^j = \mathscr{U}\left(\widehat{\mathbf{S}}_n^j, A^j\right)$; define $\widehat{\boldsymbol{\zeta}}_n(\pi)$ as the solution to $\widehat{\Lambda}_{\mathrm{ave},n}\left\{\pi, \boldsymbol{\zeta}(\pi)\right\} = 0$. As in the discounted setting, one can always find an exact root to the sample estimating equation under a linear model; however, the theory permits approximate roots as well.

To study the large sample properties of $\widehat{\boldsymbol{\zeta}}_n(\pi)$ we make use of the following regularity conditions.

(D1) There exists a non-trivial measure υ on $\mathscr{S}$ with

$$p_{\mathbf{S}^j|\mathbf{S}^{j-1},A^{j-1}}(\cdot|\mathbf{s},a) \geq \upsilon(\cdot) \text{ for all } \mathbf{s}\in\mathscr{S}, a\in\mathscr{A},$$

where $p_{\mathbf{S}^j|\mathbf{S}^{j-1},A^{j-1}}$ denotes the density of $\mathbf{S}^j$ given $\mathbf{S}^{j-1}$ and A^{j-1}.

(D2) For all $\pi \in \Pi$ and $\mathbf{s} \in \mathscr{S}$,

$$\limsup_{N\to\infty} \mathbb{E}^{\pi}\left(\frac{1}{N}\sum_{j=1}^{N} U^{j} | \mathbf{S}^{1} = \mathbf{s}\right) = \liminf_{N\to\infty} \mathbb{E}^{\pi}\left(\frac{1}{N}\sum_{j=1}^{N} U^{j} | \mathbf{S}^{1} = \mathbf{s}\right).$$

(D3) For all $\mathbf{s} \in \mathscr{S}$, $\lim_{\gamma\uparrow 1}\left[v_{\text{dis}}(\pi, \mathbf{s}) - V_{\text{ave}}(\pi)/(1-\gamma)\right] = O(1)$.

(D4) For each $\pi \in \Pi$, $\boldsymbol{\zeta}^{*}(\pi)$ solves $\Lambda_{\text{ave}}(\pi, \boldsymbol{\zeta}) = 0$, where $\boldsymbol{\zeta}^{*}(\pi)$ is an interior point of $\mathscr{Z}$, and $\mathscr{Z}$ is a compact subset of $\mathbb{R}^{\dim \zeta}$.

(D5) For each $\pi \in \Pi$, there exists a sequence of $\widehat{\boldsymbol{\zeta}}_{n}(\pi) \in \mathscr{Z}$ such that $\widehat{\Lambda}_{\text{ave},n}\{\pi, \widehat{\boldsymbol{\zeta}}_{n}(\pi)\} = o_{p}(n^{-1/2})$.

(D6) $V_{\text{ave}}(\pi)$ attains its supremum over Π at π^{*}_{ave}.

(D7) There exists a sequence $\widehat{\pi}_{\text{ave},n} \in \Pi$ such that $\widehat{V}_{\text{ave},n}(\widehat{\pi}_{\text{ave},n}) \geq \sup_{\pi\in\Pi} \widehat{V}_{\text{ave},n}(\pi) - o_{p}(1)$.

(D8) There exists a constant $c > 0$, such that

$$\mathbf{w}^{\intercal}\mathbb{E}\left\{\frac{\pi_{A^{j}}(\mathbf{S}^{j})}{P(A^{j}|\mathbf{S}^{j})}\psi_{1}(\mathbf{S}^{j})\psi_{2}(\mathbf{S}^{j}, \mathbf{S}^{j+1})^{\intercal}\right\}\mathbf{w} \geq c\|\mathbf{w}\|_{2}^{2},$$

for all $j > 0$ and $\mathbf{w} \neq \mathbf{0}$.

(D9) For each $j = 1, \ldots, J$, $\pi \in \Pi$, define

$$\mathbf{G}^{j}_{\text{ave}}\left(\pi, \mathbf{h}^{j}, a^{j}; \boldsymbol{\varrho}\right) \triangleq \frac{\pi_{a^{j}}\left\{b^{j}\left(\mathbf{h}^{j}; \boldsymbol{\varrho}\right)\right\}}{P\left\{a^{j}|b^{j}\left(\mathbf{h}^{j}; \boldsymbol{\varrho}\right)\right\}}\mathscr{U}\left\{b^{j}\left(\mathbf{h}^{j}; \boldsymbol{\varrho}\right), a^{j}\right\}\psi_{1}\left\{b^{j}\left(\mathbf{h}^{j}; \boldsymbol{\varrho}\right)\right\},$$

there exists a linear operator $\mathbf{W}^{j}_{\text{ave}}(\pi, \mathbf{h}^{j}, a^{j}; \boldsymbol{\varrho})$ such that, $\mathbb{E}\|\mathbf{W}^{j}_{\text{ave}}(\pi, \mathbf{H}^{j}, A^{j}; \boldsymbol{\varrho}^{*})\|_{F} < \infty$, and for all $\mathbf{h}^{j} \in \operatorname{dom}\mathbf{H}^{j}$ and $a^{j} \in \mathscr{A}$,

$$\mathbf{G}^{j}_{\text{ave}}\left(\pi, \mathbf{h}^{j}, a^{j}; \widehat{\boldsymbol{\varrho}}_{n}\right) - \mathbf{G}^{j}_{\text{ave}}\left(\pi, \mathbf{h}^{j}, a^{j}; \boldsymbol{\varrho}^{*}\right) = \mathbf{W}^{j}_{\text{ave}}(\pi, \mathbf{h}^{j}, a^{j}; \boldsymbol{\varrho}^{*})(\widehat{\boldsymbol{\varrho}}_{n} - \boldsymbol{\varrho}^{*}) + o_{p}(n^{-1/2}).$$

Assumption (D1) is a common recurrence condition for the generative model in the average utility MDP [see Yam75; Kur86; CC88; HL91, for variants of this assumption]. This assumption ensures that there is a nonzero transition density from any starting state to any other state under all feasible regimes. A consequence is that $V_{\text{ave}}(\pi)$ does not depend on the starting state. Assumption (D2) guarantees the existence of $V_{\text{ave}}(\pi)$ as the limit of the expected average potential utility for all $\pi \in \Pi$. Assumption (D3) requires the system dynamics be such that as $\gamma \uparrow 1$, $V_{\mathrm{dis}}(\pi, \mathbf{s})$ behaves like the expected total utility starting from $\mathbf{s}$ while $V_{\mathrm{ave}}(\pi)/(1-\gamma)$ behaves like the expected total utility averaging across initial states. The remainder of the assumptions are standard regularity assumptions for Z-estimators.

Theorem 4.5.3. *Assume (A1) - (A2), (C6) - (C7), (D1) - (D8), and that Π is finite. Then as $n \to \infty$:*

1. *For any fixed regime $\pi \in \Pi$, $\widehat{\boldsymbol{\zeta}}_n(\pi) \xrightarrow{p} \boldsymbol{\zeta}^*(\pi)$, as $n \to \infty$.*
2. *$\widehat{V}_{\mathrm{ave},n}(\widehat{\pi}_{\mathrm{ave},n}) \xrightarrow{p} V_{\mathrm{ave}}(\pi^*_{\mathrm{ave}})$ as $n \to \infty$.*

The following quantities will be used defining limiting distribution of $\widehat{\boldsymbol{\zeta}}_n(\pi)$:

$$D_1(\pi) \triangleq \mathbb{E}\left\{\sum_{j=1}^{J-1} \frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j|\mathbf{S}^j)} \psi_1(\mathbf{S}^j)\psi_2(\mathbf{S}^j, \mathbf{S}^{j+1})^{\intercal}\right\},$$

$$\widehat{D}_{1,n}(\pi) \triangleq \mathbb{P}_n\left\{\sum_{j=1}^{J-1} \frac{\pi_{A^j}\left(\widehat{\mathbf{S}}^j_n\right)}{P\left(A^j|\widehat{\mathbf{S}}^j_n\right)} \psi_1\left(\widehat{\mathbf{S}}^j_n\right)\psi_2\left(\widehat{\mathbf{S}}^j_n, \widehat{\mathbf{S}}^{j+1}_n\right)^{\intercal}\right\},$$

$$D_2(\pi, \widetilde{\pi}) \triangleq \mathbb{E}\Bigg[\sum_{j=1}^{J-1} \frac{\pi_{A^j}\left(\mathbf{S}^j\right)\widetilde{\pi}_{A^j}\left(\mathbf{S}^j\right)}{P^2\left(A^j|\mathbf{S}^j\right)}\left\{U^j - \psi_2(\mathbf{S}^j, \mathbf{S}^{j+1})^{\intercal}\boldsymbol{\zeta}(\pi)\right\}$$
$$\left\{U^j - \psi_2\left(\mathbf{S}^j, \mathbf{S}^{j+1}\right)^{\intercal}\boldsymbol{\zeta}(\widetilde{\pi})\right\}\psi_1(\mathbf{S}^j)\psi_1(\mathbf{S}^j)^{\intercal}\Bigg],$$

$$\widehat{D}_{2,n}(\pi, \widetilde{\pi}) \triangleq \mathbb{P}_n\Bigg[\sum_{j=1}^{J-1} \frac{\pi_{A^j}\left(\widehat{\mathbf{S}}^j_n\right)\widetilde{\pi}_{A^j}\left(\widehat{\mathbf{S}}^j_n\right)}{P^2\left(A^j|\widehat{\mathbf{S}}^j_n\right)}\left\{\widehat{U}^j_n - \psi_2\left(\widehat{\mathbf{S}}^j_n, \widehat{\mathbf{S}}^{j+1}_n\right)^{\intercal}\widehat{\boldsymbol{\zeta}}_n(\pi)\right\}$$
$$\left\{\widehat{U}^j_n - \psi_2\left(\widehat{\mathbf{S}}^j_n, \widehat{\mathbf{S}}^{j+1}_n\right)^{\intercal}\widehat{\boldsymbol{\zeta}}_n(\widetilde{\pi})\right\}\psi_1\left(\widehat{\mathbf{S}}^j_n\right)\psi_1\left(\widehat{\mathbf{S}}^j_n\right)^{\intercal}\Bigg],$$

$$D_3(\pi) \triangleq \mathbb{E}\left[\sum_{j=1}^{J-1} \mathbf{W}^j_{\mathrm{ave}}(\pi, \mathbf{H}^j, A^j; \boldsymbol{\varrho}^*)\right],$$

$$\widehat{D}_{3,n}(\pi) \triangleq \mathbb{P}_n\left[\sum_{j=1}^{J-1} \mathbf{W}^j_{\mathrm{ave}}(\pi, \mathbf{H}^j, A^j; \widehat{\boldsymbol{\varrho}}_n)\right].$$

Corollary 4.5.1. *Assume (A1) - (A2), (C6) - (C7), (D1) - (D9), and that Π is finite. Then for each $\pi \in \Pi$, as $n \to \infty$:*

$$\sqrt{n}\{\omega(\pi)\}^{-1/2}\{\widehat{V}_{\mathrm{ave},n}(\pi) - V_{\mathrm{ave}}(\pi)\} \rightsquigarrow \mathcal{N}(0,1),$$

where $\omega(\pi)$ is the element at entry $(1,1)$ of

$$\widehat{D}^{-1}_{1,n}(\pi)[\widehat{D}_{2,n}(\pi,\pi) + \widehat{D}_{3,n}(\pi)\widehat{\mathbf{I}}^{-1}_n(\widehat{\boldsymbol{\varrho}}_n)\widehat{D}^{\intercal}_{3,n}(\pi)]\widehat{D}^{-\intercal}_{1,n}(\pi).$$

4.5.4 Confidence intervals for value of an optimal treatment regime

The preceding results establish consistency and asymptotic normality jointly over any fixed set of regimes under the estimated MDP. We now illustrate how these results can be used to construct a confidence interval for the value of the optimal regime within a (possibly infinite) class of regimes. We present only the discounted utility case as the approach for the average utility is essentially the same. The strategy we follow here is in the same spirit as projection confidence intervals, which are commonly used for non-smooth functionals [see BB94; Rob04; Lab14]. Let $\eta \in (0,1)$ be arbitrary. An overview of the basic approach is as follows: (S1) specify a parametric class of regimes; (S2) construct a $(1-\eta)\times 100\%$ confidence region for the parameters indexing the optimal regime; mapping each element in this region to its corresponding regime thus defines a confidence region in the space of regimes; (S3) for each regime in the confidence region, construct a $(1-\eta)\times 100\%$ confidence interval for its value using the asymptotic normality of the estimated value for a fixed regime (derived in the previous section); and (S4) take a union of all the intervals in the preceding step. It is easily shown that if the region constructed in (S2) is a valid confidence region and each interval in (S3) is also (marginally) valid then the union is a valid $(1-2\eta)\times 100\%$ interval for the optimal value. For additional discussion see Tsiatis et al. [Tsi19].

Our goal is to derive a confidence interval for the optimal value, $\sup_{\pi\in\Pi} V^R_{\text{dis}}(\pi)$, when Π is a parametric class of regimes. We make the following assumptions.

(C9) The class of regimes $\Pi = \{\pi(\cdot;\xi) : \xi \in \Xi\}$ is indexed by ξ, where Ξ is a compact subset of $\mathbb{R}^{\dim \xi}$.

(C10) The map $\xi \mapsto V^R_{\text{dis}}(\xi) \triangleq \int \phi^\intercal(\mathbf{s})\boldsymbol{\alpha}^*(\xi)dR(\mathbf{s})$ has a unique and well-separated maximum at $\xi = \xi^*$, which is an interior point of Ξ.

(C11) There exists a sequence of $\widehat{\xi}_n \in \Xi$ such that $\widehat{V}^R_{\text{dis},n}(\widehat{\xi}_n) \ge \sup_{\xi\in\Xi} \widehat{V}^R_{\text{dis},n}(\xi)$.

(C12) The map $\xi \mapsto \int \phi^\intercal(\mathbf{s})\boldsymbol{\alpha}^*(\xi)dR(\mathbf{s})$ is twice continuously differentiable in a neighborhood of ξ^* and $\frac{\partial^2}{\partial \xi^2}\int \phi^\intercal(\mathbf{s})\boldsymbol{\alpha}^*(\xi)dR(\mathbf{s})\Big|_{\xi=\xi^*}$ is positive definite.

Corollary 4.5.2. *Assume (A1) - (A2), (C1) - (C12). Then as $n\to\infty$:*

1. *the results in Theorem 4.5.2 hold over all $\pi \in \Pi$;*

2. $\sqrt{n}(\widehat{\xi}_n - \xi^*) \rightsquigarrow \mathcal{N}(\mathbf{0}, \circ_\xi)$, *where* $\circ_\xi = \circ_{1,\xi}^{-1}\circ_{2,\xi}\circ_{1,\xi}^{-1}$,

$$\circ_{1,\xi} = \int \frac{\partial^2}{\partial \xi^2}\phi^\intercal(\mathbf{s})\boldsymbol{\alpha}^*(\xi)dR(\mathbf{s})\Big|_{\xi=\xi^*},$$

$$\circ_{2,\xi} = \int \left\{ \frac{\partial}{\partial \xi} \phi^{\intercal}(\mathbf{s}) \boldsymbol{\alpha}^{*}(\xi) \right\} \left\{ \frac{\partial}{\partial \xi} \phi^{\intercal}(\mathbf{s}) \boldsymbol{\alpha}^{*}(\xi) \right\}^{\intercal} dR(\mathbf{s}) \Big|_{\xi=\xi^{*}}.$$

Corollary 4.5.3. *Assume (A1) - (A2), (C1) - (C12). Let* $\eta \in (0,1)$ *be arbitrary. Define*

$$\mathscr{E}_{1-\eta,n} \triangleq \{\xi : n(\widehat{\xi}_n - \xi)^{\intercal} \widehat{\circ}_{\xi,n} (\widehat{\xi}_n - \xi) \leq \chi^2_{\dim \xi, 1-\eta}\},$$

where $\widehat{\circ}_{\xi,n}$ *is the sample analog of* $\circ_{\xi}$. *Let* $z_{\eta/2}(\xi)$ *and* $z_{1-\eta/2}(\xi)$ *be the* $(\eta/2) \times 100$ *and* $(1-\eta/2) \times 100$ *percentiles of a Gaussian distribution with mean-zero and variance* $\widehat{\sigma}^{-1/2}_{\mathrm{dis},n}(\xi)$. *Then it follows that*

$$P\left[\inf_{\xi \in \mathscr{E}_{1-\eta,n}} \left\{ \frac{z_{\eta/2}(\xi)}{\sqrt{n}} + \widehat{V}_{\mathrm{dis},n}(\xi) \right\} \leq V^{R}_{\mathrm{dis}}(\xi^{*}) \leq \sup_{\xi \in \mathscr{E}_{1-\eta,n}} \left\{ \frac{z_{1-\eta/2}(\xi)}{\sqrt{n}} + \widehat{V}_{\mathrm{dis},n}(\xi) \right\} \right] \geq 1 - 2\eta.$$

While projection intervals can be extremely conservative in some settings [see Lab14], in our simulation experiments, which are based on the STEP-BD study data, the degree of conservatism was relatively mild. Thus, these intervals appear to be suitable for application with data like STEP-BD.

4.6 Simulation experiments

We study the finite sample performance of the proposed point and interval estimators using a series of simulation experiments. The data-generating models we consider are designed to mimic salient features of the STEP-BD trial. We consider a follow-up period of one year. At each visit, $j \geq 1$, we observe three patient covariates, $\mathbf{X}^j \in \mathbb{R}^3$, and a treatment is chosen from among three candidates $A^j \in \{1,2,3\}$ so that

$$\begin{aligned}
\mathrm{logit}\left\{ \frac{P(A^j = 1)}{P(A^j = 3)} \right\} &= -0.2 + 0.1X_1^j - 0.1X_2^j + 0.1X_3^j, \\
\mathrm{logit}\left\{ \frac{P(A^j = 2)}{P(A^j = 3)} \right\} &= -0.2 - 0.1X_1^j + 0.1X_2^j - 0.1X_3^j.
\end{aligned}$$

We consider five latent states intended to encode the health states: depression, mania, mixed type, hypomania, and stable; thus, $\mathbf{B}^j$ is an element of the five-dimensional probability simplex. The j^{th} interarrival time between visits follows an exponential distribution with rate $e^{\{}e_1 + 0.1(B_1^j + B_2^j) - 0.1(B_3^j + B_4^j)\}$, where $e_1 \overset{i.i.d.}{\sim}$ Uniform$(-3,-2)$ is a subject-specific random effect. We assume that the conditional distribution of $\mathbf{X}^j$ given $M^j, \mathbf{X}^{j-1}$ follows a Gaussian autoregressive model (see Section 4 for the form of the density) indexed by the following parameters:

$$
\begin{array}{llll}
\text{State 1}: & \boldsymbol{\mu}_1 = (2,2,2)^\intercal & \Psi_1 = \frac{1}{10}\mathbf{I}_{3\times3} & \Sigma_1 = \frac{1}{10}\mathbf{I}_{3\times3} + \frac{1}{10}\mathbf{1}_{3\times3}, \\
\text{State 2}: & \boldsymbol{\mu}_2 = (2,1,-2)^\intercal & \Psi_2 = \frac{1}{10}\mathbf{I}_{3\times3} & \Sigma_2 = \frac{1}{10}\mathbf{I}_{3\times3} + \frac{1}{10}\mathbf{1}_{3\times3}, \\
\text{State 3}: & \boldsymbol{\mu}_3 = (-2,1,2)^\intercal & \Psi_3 = -\frac{1}{10}\mathbf{I}_{3\times3} & \Sigma_3 = \frac{3}{10}\mathbf{I}_{3\times3} - \frac{1}{10}\mathbf{1}_{3\times3}, \\
\text{State 4}: & \boldsymbol{\mu}_4 = (-2,-2,-2)^\intercal & \Psi_4 = -\frac{1}{10}\mathbf{I}_{3\times3} & \Sigma_4 = \frac{3}{10}\mathbf{I}_{3\times3} - \frac{1}{10}\mathbf{1}_{3\times3}, \\
\text{State 5}: & \boldsymbol{\mu}_5 = (0,0,0)^\intercal & \Psi_5 = \mathbf{0}_{3\times3} & \Sigma_5 = \mathbf{I}_{3\times3},
\end{array}
$$

where $\boldsymbol{\mu}_k$, Σ_k, and Ψ_k, $k = 1,2,\ldots,5$, are state-dependent mean, covariance, and autoregression coefficients; $\mathbf{0}_{3\times3}$ is a 3-by-3 matrix of zeros and $\mathbf{1}_{3\times3}$ is a 3-by-3 matrix of ones.

In the first scenario, we consider the case where the evolution of latent disease status follows a first-order Markov process, i.e., the generative model is correctly specified. We consider the following utility function

$$
U^j = 2 - |X_1^{j+1}| - |X_3^{j+1}|;
$$

thus the utility is larger when X_1 and X_3 are close to 0. In these simulation experiments, we might think of X_1 and X_3 as symptom severity measures represented as deviations from a stable condition (coded as zero). The off-diagonals in the transition rate matrix $\{q_{m,\ell}(a)\}_{m\neq\ell}$ are

$$
\begin{array}{ll}
\operatorname{logit} q_{k,\ell} = e_3 + 5\cdot I(A=1) & \text{for } (k,\ell)\in\{(1,5),(4,5),(2,3),(3,2)\}, \\
\operatorname{logit} q_{k,\ell} = e_3 + 5\cdot I(A=2) & \text{for } (k,\ell)\in\{(2,5),(3,5),(1,4),(4,1)\}, \\
\operatorname{logit} q_{k,\ell} = e_3 + 2\cdot I(A=1) + 2\cdot I(A=2) & \text{for } (k,\ell)\in\{(5,1),(5,2),(5,3),(5,4)\}, \\
\operatorname{logit} q_{k,\ell} = e_3 & \text{otherwise},
\end{array}
$$

where $e_3 \overset{i.i.d}{\sim} \text{Uniform}(-7,-6)$ are subject-specific random effects. The diagonals are thus $q_{k,k} = -\sum_{\ell\neq k} q_{k,\ell}$ for $k = 1,\ldots,5$. In this setup, treatment 1 will: (1) increase the probability of transitioning to state 5 when the current state is either 1 or 4; (2) increase the probability of transitioning between state 2 and 3; (3) and increase the probability of transitioning out of state 5.

In the second scenario, we consider the case where the generative model is misspecified. At visit j, the latent disease states are distributed according to a multinomial distribution with parameters $(p_1, p_2, p_3, p_4, p_5)$, which are drawn from a Dirichlet distribution with parameter $(1,1,1,1,1)$ so that the latent disease state distribution is randomly drawn at each visit. We consider the utility function of the form

$$U^j = (B_1^j + B_4^j)\{2I(A^j = 1) - 1\} + (B_2^j + B_3^j)\{2I(A^j = 2) - 1\} + B_5^j\{2I(A^j = 3) - 1\},$$

which indicates: treatment 1 is the optimal treatment when the current state is either 1 or 4; treatment 2 is the optimal treatment when the current state is either 2 or 3; and treatment 3 is the optimal treatment when the current state is 5. Because this utility is not directly observed, the estimated optimal regime is constructed with the *estimated* utility; however, evaluations are made and reported for the true utility.

We evaluate the mean and standard error of the value (marginal mean outcome) of candidate regimes under sample sizes 100 and 200. We consider both stochastic and deterministic regimes; when estimating stochastic regimes we used an L_2 penalty tuned to ensure that each treatment is selected with (estimated) probability at least 0.05 across all observed states. The stochastic regimes we consider include the data-generating regime, the proposed POMDP regimes in the form of multinomial logistic regression using both linear and quadratic basis functions, and their MDP regime counterparts, which do not utilize latent state information. The deterministic regimes we consider include the optimal regime, the proposed POMDP regimes with linear policies indexed by linear and quadratic basis functions, and their MDP regime counterparts, which do not use latent state information. All results are based on 500 Monte Carlo replications.

Table 4.1 shows the mean and standard error for the estimated values in scenario 1, where the generative model is correctly specified. The proposed POMDP estimators outperform the baseline MDP estimators in both deterministic and stochastic regimes. The results are consistent in both the discounted reward and average reward settings. The POMDP estimators have higher mean values and smaller standard errors than their MDP counterparts. Indeed, the values from the estimated POMDP regimes are close to those of the true optimal deterministic regime. In Table 4.2, where the POMDP model is misspecified, the estimated values from the POMDP regimes still significantly outperform the observed and MDP regimes. This result suggests that the linear model may be robust to moderate misspecification. The inclusion of quadratic terms did not greatly affect performance. Table 4.3 shows the coverage probability and half width of the proposed confidence interval for the optimal value under linear regimes when the model is correctly specified. The confidence intervals attain nominal (95%) coverage, although they are a bit conservative as expected.

Table 4.1 Mean (standard error) for the estimated values for stochastic and deterministic regimes in scenario 1.

	Stochastic regimes					Deterministic regimes				
n	V_{obs}	$V_{\text{lin}}^{\text{MDP}}$	$V_{\text{quad}}^{\text{MDP}}$	$V_{\text{lin}}^{\text{POM}}$	$V_{\text{quad}}^{\text{POM}}$	$V_{\text{lin}}^{\text{MDP}}$	$V_{\text{quad}}^{\text{MDP}}$	$V_{\text{lin}}^{\text{POM}}$	$V_{\text{quad}}^{\text{POM}}$	V^{opt}
Discounted reward										
100	-4.602	-5.978	-5.888	1.015	1.037	-5.878	-5.886	1.213	1.158	1.144
	(.684)	(3.517)	(3.465)	(.882)	(.851)	(3.527)	(3.508)	(1.060)	(1.054)	(1.076)
200	-4.622	-5.321	-5.355	1.106	1.062	-5.390	-5.422	1.265	1.248	1.285
	(.527)	(3.273)	(3.279)	(.584)	(.606)	(3.393)	(3.359)	(.701)	(.703)	(.689)
Average reward										
100	-0.481	-0.951	-0.954	0.032	0.014	-0.892	-0.893	-0.079	-0.085	-0.063
	(.076)	(.494)	(.500)	(.169)	(.164)	(.448)	(.445)	(.186)	(.181)	(.218)
200	-0.483	-0.881	-0.882	0.053	0.051	-0.808	-0.807	-0.066	-0.066	-0.017
	(.052)	(.516)	(.518)	(.110)	(.116)	(.444)	(.442)	(.144)	(.135)	(.122)

Table 4.2 Mean (standard error) for the estimated values for stochastic and deterministic regimes in scenario 2.

	Stochastic regimes					Deterministic regimes				
n	V_{obs}	$V_{\text{lin}}^{\text{MDP}}$	$V_{\text{quad}}^{\text{MDP}}$	$V_{\text{lin}}^{\text{POM}}$	$V_{\text{quad}}^{\text{POM}}$	$V_{\text{lin}}^{\text{MDP}}$	$V_{\text{quad}}^{\text{MDP}}$	$V_{\text{lin}}^{\text{POM}}$	$V_{\text{quad}}^{\text{POM}}$	V^{opt}
Discounted reward										
100	-2.921	0.412	0.399	3.007	3.009	0.478	0.466	3.660	3.657	3.804
	(.168)	(.152)	(.154)	(.153)	(.156)	(.154)	(.157)	(.162)	(.164)	(.091)
200	-2.909	0.420	0.420	3.026	3.024	0.482	0.491	3.671	3.672	3.807
	(.112)	(.107)	(.103)	(.085)	(.083)	(.101)	(.103)	(.076)	(.078)	(.066)
Average reward										
100	-0.343	0.021	0.021	0.262	0.263	0.033	0.031	0.368	0.370	0.418
	(.020)	(.026)	(.024)	(.046)	(.048)	(.025)	(.026)	(.074)	(.074)	(.011)
200	-0.342	0.024	0.024	0.270	0.271	0.033	0.034	0.384	0.384	0.418
	(.014)	(.019)	(.019)	(.036)	(.035)	(.020)	(.021)	(.057)	(.057)	(.008)

Table 4.3 Coverage probability and half width of the confidence intervals at 0.05 nominal level for the linear POMDP regimes when the model is correctly specified.

		Coverage probability		Half width	
Criterion	n	Stochastic	Deterministic	Stochastic	Deterministic
Discounted	100	0.986	0.986	1.378	1.451
Discounted	200	0.988	0.984	0.966	1.010
Average	100	0.982	0.980	0.238	0.225
Average	200	0.978	0.990	0.184	0.185

4.7 Case study

The data used in our case study are derived from the standard care pathway of the STEP-BD clinical trial [Sac03]. Inclusion criteria required that patients be (i) at least 18 years old and (ii) diagnosed with bipolar type I or bipolar type II disorder at screening. Treatment decisions at each clinic visit were made based on doctor-patient preference and thus the data are observational. Validity of the proposed methods thus requires additional causal assumptions. As these assumptions are standard, we have relegated them to the technical details at the end of the chapter.

Figure 4.1 shows the treatment histories for a sample of patients in the STEP-BD observational pathway. It can be seen that the timing, number, type, and dosage of treatment varies widely across patients.

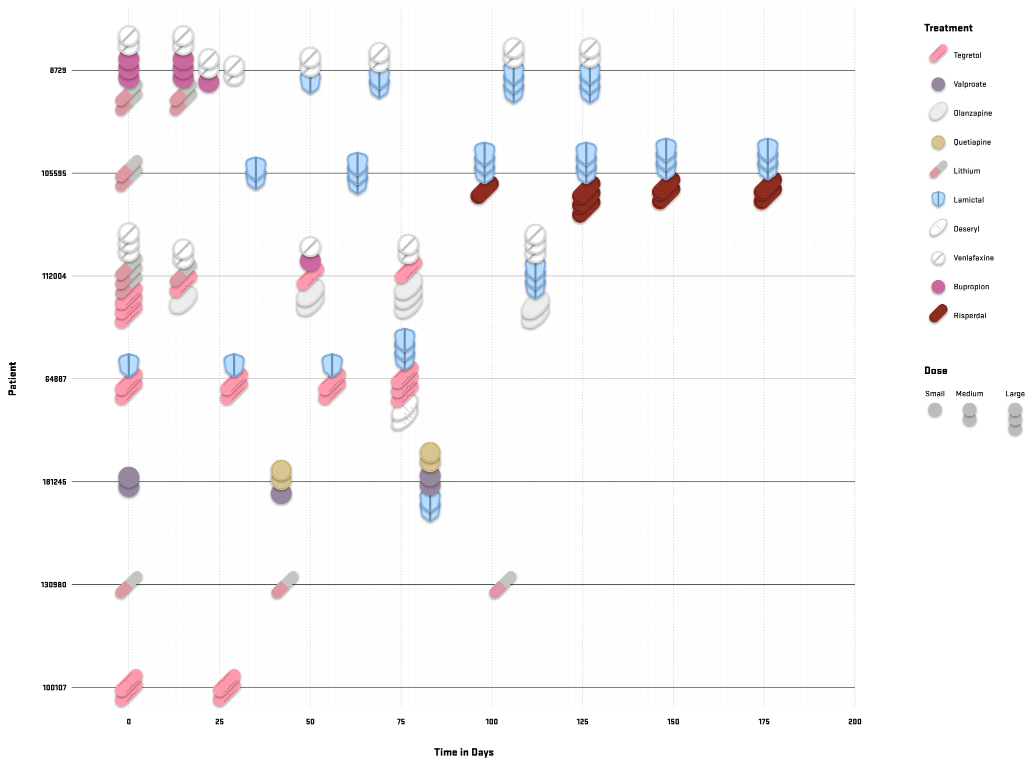

Figure 4.1 Treatment histories for a sample of patients in the STEP-BD observational pathway. The timing, number, type, and dosage of treatment varies widely across patients.

We group the medications into two groups: (A) antidepressants and (M) mood stabilizers, the dose levels of which are divided into three levels: low, medium, and high. The categorization of antidepressants and mood stabilizers as well as the corresponding dose levels are provided in the Appendix. Table 4.4 enumerates the 15 potential treatment combinations.

Table 4.4 List of potential treatment combinations. A = antidepressants, M = mood stabilizers.

Treatment ID	Treatment combinations
1	low A
2	medium A
3	high A
4	low M
5	medium M
6	high M
7	low A + low M
8	low A + medium M
9	low A + high M
10	medium A + low M
11	medium A + medium M
12	medium A + high M
13	high A + low M
14	high A + medium M
15	high A + high M

We assume that there are five latent states corresponding to: depression, mania, mixed type, hypomania, and stable moods. At each stage, a patient's state is constructed using the latent state probability vector and observable patient covariates: age, bipolar disorder type, sum of depression score (SUMD), sum of mania score (SUMM), percent of days depressed, percent of days low interest in most activities, and percent of days with abnormal mood elevation. SUMD and SUMM are aggregates of multiple items on a questionnaire. In the study protocol, both SUMD and SUMM are defined to be missing when an answer to any of the inventory questions is missing, which results in 19% and 36% missing entries, respectively. Multiple imputation [Rub04] is used for missing items, and aggregated scores are recalculated using the imputed data. Besides the inventory questions in SUMD and SUMM, other variables used in multiple imputation include the three other continuous patient covariates as well as patient baseline characteristics (details and code are provided in Supplemental Materials). We imputed five complete data sets, and to each imputed data set we applied the proposed methodology to estimate the optimal treatment regime. Parameters indexing each estimated optimal treatment regime were then averaged and used in the final estimated optimal treatment regime. The utility at each stage is defined as $2-\text{SUMD}-\text{SUMM}$, where both SUMD and SUMM are standardized to fall between 0 and 1. A higher utility implies a lower SUMD and SUMM, which corresponds to a more desirable clinical outcome. [Wu15] used SUMD as the clinical outcome in estimating the optimal treatment regime in the randomized arm of STEP-BD, where there were only two decision stages. However, an effective long-term treatment regime for bipolar disorder should alleviate depression symptoms without inducing mania episodes, which is why we opted for a composite outcome.

Table 4.5 shows the mean and standard error of the estimated mood state probabilities using the proposed latent Markov model. The results are promising in that they largely agree with the reported clinical status on the clinical monitoring form. The model had some difficulty delineating between mania and hypomania; however, this is not surprising as (clinically) these abnormal states differ only in severity.

Table 4.6 shows the value under the observed regime and under the estimated regime for average utility and discounted utility ($\gamma = 0.95$) setup. In both cases, the lower bound of the 95% confidence interval for the value of the estimated regime is higher than the observed regime.

Figure 4.2 shows the estimated optimal treatment regime obtained by maximizing the average utility, which is projected onto a decision tree for ease of interpretation. The predicted optimal treatment is either mood stabilizers or a combination of mood stabilizers and antidepressants, i.e., it never recommends antidepressants alone. Such recommendations are anticipated by the clinical belief that prescribing antidepressants alone for bipolar disorder patients may increase the risk of inducing a manic episode [Pat15]. The estimated optimal regime also prescribes antidepressants only as a supplement for the mood stabilizer either when there is some evidence of depression (SUMD is large or probability of depression is large) or there is little evidence of mania (SUMM is small or probability of mania is small). The estimated optimal treatment regime for the discounted utility ($\gamma = 0.95$) is included in the Appendix and is qualitatively similar.

Table 4.5 Mean estimated mood state probabilities within each of the five clinical status categories. The standard errors range from 0.0001 to 0.01.

Clinical status	$\widehat{P}$(Depress)	$\widehat{P}$(Mania)	$\widehat{P}$(Mixed)	$\widehat{P}$(Hypomania)	$\widehat{P}$(Stable)
Depression	0.85	<0.01	0.14	<0.01	0.05
Mania	<0.01	0.55	0.05	0.47	<0.01
Mixed	0.13	0.09	0.84	0.08	<0.01
Hypomania	<0.01	0.34	0.09	0.44	0.02
Stable	0.02	<0.01	<0.01	<0.01	0.92

Table 4.6 Comparison of the value under the observed regime and the value under the estimated regime for average utility and discounted utility ($\gamma = 0.95$).

Criterion	Observed value	Value under estimated regime (95% C.I.)
Average utility	1.66	1.81 (1.75, 1.87)
Discounted utility	14.40	33.76 (20.66, 45.71)

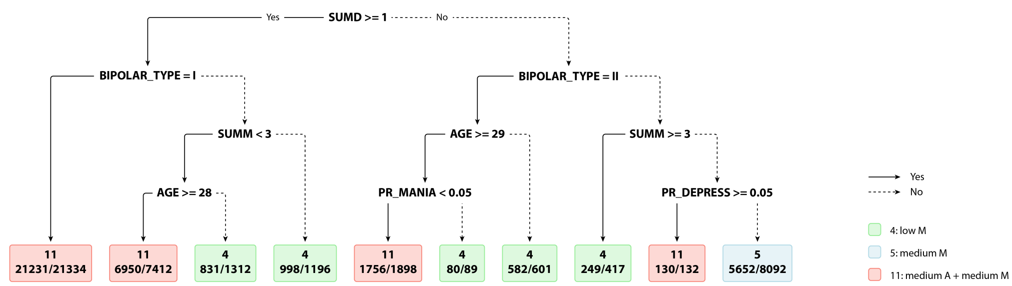

Figure 4.2 Estimated optimal treatment regime obtained by maximizing average utility, which is projected onto a decision tree for interpretation. Each tree node shows the splitting criterion, majority treatment label, and its proportion.

4.8 Conclusions

We developed a framework for estimation of an optimal treatment regime using data from long-term observational or randomized clinical studies. A key contribution of this work is incorporation of a patient's latent health status, e.g., their true mood state in the context of bipolar depression. We showed that using this structure can lead to estimated optimal regimes that are clinically meaningful and that significantly outperform methods that fail to use this structure.

CHAPTER

5

CONCLUSIONS

In Chapter 2, a two-stage subject-specific hidden semi-Markov model is proposed to model longitudinal, intensive accelerometer data with excessive proportions of zeros. An alternative to the proposed two-stage fitting procedure is to use a mixed hidden Markov model [Alt07]. The primary advantage of the proposed approach is its computational tractability and efficiency. The fitting of a mixed hidden Markov model involves integration of multiple nested random effects in each hidden state, which makes computing the log-likelihood prohibitively burdensome. Our refinement of the analysis of high-volume activity data in this naturally occurring model of pain provides a clinically relevant outcome measure that can be used in naturally occurring models of arthritis in companion animals to provide highly relevant data on the potential efficacy of putative drugs prior to Phase II and III efficacy studies in humans, thus improving the translational paradigm by optimizing the critical go/no-go decision prior to Phase II. Such sophisticated approaches to high-volume activity data have not been applied to humans and our approach also has relevance to the analysis of activity data in humans to better understand the impact of therapeutics in a more refined manner.

In Chapter 3, we used a hierarchical continuous-time hidden Markov model to estimate the human physical activity profile measured by wearable monitors. In the future, this framework can be further extended to the controlled clinical studies to estimate certain treatment effects in a specific cohort of patients. We can also allow for time-varying covariates in the transition rates. Moreover, when some model parameters are truly subject-specific or subgroup-specific, it may be more powerful to model them as random so that tests based on variance components can be constructed to test their effects. Another modification is to extend the latent continuous-time Markov process to a semi-Markov process. This will be scientifically interesting because it is reasonable to assume that the current latent state not only depends on the most recent past state but also on the history of the state trajectory. However, all these changes are computationally expensive, especially on such large-scale high-frequency data. Corresponding estimation methods have to be developed before the application becomes feasible.

In Chapter 4, we propose a method for estimating an optimal treatment regime in the indefinite-time-horizon setting when data are irregularly spaced, contain multiple treatment changes, and cannot be assumed to be Markov. A key contribution of this work is incorporation of a patient's latent health status, e.g., their true mood state in the context of bipolar depression. Other contributions of this work include: a theory-driven construction of the latent-process model, the application of POMDPs to episodic chronic mental illness, and the development valid statistical inference for clinically relevant estimands in this context. The proposed methodology is extensible and could be ported for estimation and inference with optimal treatment regimes in other contexts having complex treatment and observation patterns, e.g., mobile-health. An alternative view of this work (divorced from any clinical impact) is as a method for transforming non-Markov decision processes so that they are approximately homogeneous and Markov and thus amenable to analyses using existing tools for MDPs. In this line, one could generalize the proposed methodology to include continuous latent processes. Such an approach is aligned with existing work in POMDPs in the computer science and engineering literature. We leave such extensions to future work.

BIBLIOGRAPHY

[Alb62] Albert, A. "Estimating the infinitesimal generator of a continuous time, finite state Markov process". *The Annals of Mathematical Statistics* (1962), pp. 727–753.

[All09] Allman, E. S. et al. "Identifiability of parameters in latent structure models with many observed variables". *The Annals of Statistics* **37**.6A (2009), pp. 3099–3132.

[Alt07] Altman, R. M. "Mixed hidden Markov models: an extension of the hidden Markov model to the longitudinal data setting". *Journal of the American Statistical Association* **102**.477 (2007), pp. 201–210.

[An13] An, Y. et al. *Identifiability and inference of hidden Markov models*. Tech. rep. Technical report, 2013.

[Ast65] Astrom, K. J. "Optimal control of Markov decision processes with incomplete state estimation". *Journal of Mathematical Analysis and Applications* **10** (1965), pp. 174 –205.

[AI16] Athey, S. & Imbens, G. "Recursive partitioning for heterogeneous causal effects". *Proceedings of the National Academy of Sciences* **113**.27 (2016), pp. 7353–7360.

[AL06] Athreya, K. B. & Lahiri, S. N. *Measure theory and probability theory*. Springer Science & Business Media, 2006.

[Bai14] Bai, A. et al. "Thompson sampling based Monte-Carlo planning in POMDPs". *Twenty-Fourth International Conference on Automated Planning and Scheduling*. 2014.

[BL09] Barbu, V. S. & Limnios, N. *Semi-Markov chains and hidden semi-Markov models toward applications: their use in reliability and DNA analysis*. Vol. 191. Springer Science & Business Media, 2009.

[Bar05] Bartalesi, R et al. "Wearable kinesthetic system for capturing and classifying upper limb gesture". *Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2005. World Haptics 2005. First Joint*. IEEE. 2005, pp. 535–536.

[BP66] Baum, L. E. & Petrie, T. "Statistical inference for probabilistic functions of finite state Markov chains". *The annals of mathematical statistics* **37**.6 (1966), pp. 1554–1563.

[Bau70] Baum, L. E. et al. "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains". *The annals of mathematical statistics* **41**.1 (1970), pp. 164–171.

[BB94] Berger, R. L. & Boos, D. D. "P values maximized over a confidence set for the nuisance parameter". *Journal of the American Statistical Association* **89**.427 (1994), pp. 1012–1016.

[Bic98] Bickel, P. J. et al. "Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models". *The Annals of Statistics* **26**.4 (1998), pp. 1614–1635.

[Bon11] Bonsall, M. B. et al. "Nonlinear time-series approaches in characterizing mood stability and mood instability in bipolar disorder". *Proceedings of the Royal Society B: Biological Sciences* **279**.1730 (2011), pp. 916–924.

[Bon15] Bonsall, M. B. et al. "Bipolar disorder dynamics: affective instabilities, relaxation oscillations and noise". *Journal of the Royal Society Interface* **12**.112 (2015), p. 20150670.

[Boy11] Boyd, S. et al. "Distributed optimization and statistical learning via the alternating direction method of multipliers". *Foundations and Trends® in Machine Learning* **3**.1 (2011), pp. 1–122.

[Buc04] Buckwalter, J. A. et al. "The impact of osteoarthritis: implications for research." *Clinical orthopaedics and related research* **427** (2004), S6–S15.

[BB06] Bulla, J. & Bulla, I. "Stylized facts of financial time series and hidden semi-Markov models". *Computational Statistics & Data Analysis* **51**.4 (2006), pp. 2192–2209.

[Bur03] Bureau, A. et al. "Applications of continuous time hidden Markov models to the study of misclassified disease outcomes". *Statistics in medicine* **22**.3 (2003), pp. 441–462.

[Bus17] Busoniu, L. et al. *Reinforcement learning and dynamic programming using function approximators*. CRC press, 2017.

[Cap05] Cappé, O. et al. "Inference in hidden markov models". *Springer Series in Statistics*. 2005.

[Cap09] Cappé, O. et al. "Inference in hidden markov models". *Proceedings of EUSFLAT Conference*. 2009, pp. 14–16.

[Cas98] Cassandra, A. R. "A survey of POMDP applications". *Working notes of AAAI 1998 fall symposium on planning with partially observable Markov decision processes*. Vol. 1724. 1998.

[Cas13] Cassandra, A. R. et al. “Incremental pruning: A simple, fast, exact method for partially observable Markov decision processes”. *arXiv preprint arXiv:1302.1525* (2013).

[Cat05] Catellier, D. J. et al. “Imputation of missing data when measuring physical activity by accelerometry”. *Medicine and science in sports and exercise* **37**.11 Suppl (2005), S555.

[CC88] Cavazos-Cadena, R. “Necessary and sufficient conditions for a bounded solution to the optimality equation in average reward Markov decision chains”. *Systems & control letters* **10**.1 (1988), pp. 71–78.

[CM14] Chakraborty, B. & Murphy, S. A. “Dynamic treatment regimes”. *Annual review of statistics and its application* **1** (2014), pp. 447–464.

[Che12] Chen, A. et al. “The global economic cost of osteoarthritis: how the UK compares”. *Arthritis* **2012** (2012).

[Che88] Cheng, H.-T. “Algorithms for partially observable Markov decision processes”. PhD thesis. University of British Columbia, 1988.

[Cia15] Ciarleglio, A. et al. “Treatment decisions based on scalar and functional baseline covariates”. *Biometrics* **71**.4 (2015), pp. 884–894.

[Cia16] Ciarleglio, A. et al. “Flexible functional regression methods for estimating individualized treatment rules”. *Stat* **5**.1 (2016), pp. 185–199.

[Cou11] Council, N. R. et al. “Relieving Pain in America: A Blueprint for Transforming Prevention, Care”. *Education, and Research. Washington, DC: The National Academies Press* (2011).

[Cra04] Cradock, A. L. et al. “Youth recall and TriTrac accelerometer estimates of physical activity levels.” *Medicine and science in sports and exercise* **36**.3 (2004), pp. 525–532.

[Dau09] Daugherty, D. et al. “Mathematical models of bipolar disorder”. *Communications in Nonlinear Science and Numerical Simulation* **14**.7 (2009), pp. 2897–2908.

[DB11] DeSantis, S. M. & Bandyopadhyay, D. “Hidden Markov models for zero-inflated Poisson counts with an application to substance use”. *Statistics in medicine* **30**.14 (2011), pp. 1678–1694.

[Dig02] Diggle, P. *Analysis of longitudinal data*. Oxford University Press, 2002.

[DM01] Douc, R., Matias, C., et al. “Asymptotics of the maximum likelihood estimator for general hidden Markov models”. *Bernoulli* **7**.3 (2001), pp. 381–420.

[Dou04] Douc, R. et al. “Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime”. *The Annals of statistics* **32**.5 (2004), pp. 2254–2304.

[Dra62] Drake, A. W. “Observation of a Markov process through a noisy channel”. PhD thesis. Massachusetts Institute of Technology, 1962.

[Dra67] Drake, A. W. *Fundamentals of Applied Probability Theory*. Mcgraw-Hill College, 1967.

[Efr92] Efron, B. “Bootstrap methods: another look at the jackknife”. *Breakthroughs in statistics*. Springer, 1992, pp. 569–593.

[Eld00] Elders, M. J. “The increasing impact of arthritis on public health.” *The Journal of Rheumatology. Supplement* **60** (2000), pp. 6–8.

[Ert19] Ertefaie, A. “Constructing dynamic treatment regimes in infinite-horizon settings”. *arXiv preprint arXiv:1406.0764* (2019).

[Eve11] Evenson, K. R. “Towards an understanding of change in physical activity from pregnancy through postpartum”. *Psychology of sport and exercise* **12**.1 (2011), pp. 36–45.

[Fed19] Fedus, W. et al. “Hyperbolic discounting and learning over multiple horizons”. *arXiv preprint arXiv:1902.06865* (2019).

[Fer80] Ferguson, J. D. “Variable duration models for speech”. *Proceedings of the Symposium on the Application of HMMs to Text and Speech*. 1980, pp. 143–179.

[Gae09] Gael, J. V. et al. “The infinite factorial hidden Markov model”. *Advances in Neural Information Processing Systems*. 2009, pp. 1697–1704.

[Gas13] Gassiat, E. et al. “Finite state space non parametric Hidden Markov Models are in general identifiable”. *arXiv preprint arXiv:1306.4657* (2013).

[GJ96] Ghahramani, Z. & Jordan, M. I. “Factorial hidden Markov models”. *Advances in Neural Information Processing Systems*. 1996, pp. 472–478.

[Gru15] Gruen, M. E. et al. “Criterion validation testing of clinical metrology instruments for measuring degenerative joint disease associated mobility impairment in cats”. *PLoS One* **10**.7 (2015), e0131839.

[Gru17] Gruen, M. E. et al. “The Use of Functional Data Analysis to Evaluate Activity in a Spontaneous Model of Degenerative Joint Disease Associated Pain in Cats”. *PloS one* **12**.1 (2017), e0169576.

[Gru14] Gruen, M. et al. "Detection of clinically relevant pain relief in cats with degenerative joint disease associated pain". *Journal of veterinary internal medicine* **28**.2 (2014), pp. 346–350.

[Gru16] Gruen, M. et al. "A Feline-Specific Anti-Nerve Growth Factor Antibody Improves Mobility in Cats with Degenerative Joint Disease–Associated Pain: A Pilot Proof of Concept Study". *Journal of veterinary internal medicine* **30**.4 (2016), pp. 1138–1148.

[Had85] Hadler, N. "Osteoarthritis as a public health problem." *Clinics in rheumatic diseases* **11**.2 (1985), p. 175.

[Ham89] Hamilton, J. D. "A new approach to the economic analysis of nonstationary time series and the business cycle". *Econometrica: Journal of the Econometric Society* (1989), pp. 357–384.

[Ham96] Hamilton, J. D. "Specification testing in Markov-switching time-series models". *Journal of econometrics* **70**.1 (1996), pp. 127–157.

[Han18] Hanna, J. P. et al. "Importance sampling policy evaluation with an estimated behavior policy". *arXiv preprint arXiv:1806.01347* (2018).

[Han12] Hansen, B. H. et al. "Accelerometer-determined physical activity in adults and older people." *Medicine and science in sports and exercise* **44**.2 (2012), pp. 266–272.

[Han98a] Hansen, E. A. "An improved policy iteration algorithm for partially observable MDPs". *Advances in Neural Information Processing Systems*. 1998, pp. 1015–1021.

[Han98b] Hansen, E. A. "Solving POMDPs by searching in policy space". *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc. 1998, pp. 211–219.

[Han74] Hanushek, E. A. "Efficient estimators for regressing regression coefficients". *The American Statistician* **28**.2 (1974), pp. 66–67.

[Hay14] Hayes, A. G. et al. "Multiple mechanisms have been tested in pain—how can we improve the chances of success?" *Current opinion in pharmacology* **14** (2014), pp. 11–17.

[He07] He, J. et al. "Real-time daily activity classification with wireless sensor networks using hidden markov model". *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. IEEE. 2007, pp. 3192–3195.

[Hen10] Henderson, R. et al. “Regret-regression for optimal dynamic treatment regimes”. *Biometrics* **66**.4 (2010), pp. 1192–1201.

[HL91] Hernández-Lerma, O. et al. “Recurrence conditions for Markov decision processes with Borel state space: a survey”. *Annals of Operations Research* **28**.1 (1991), pp. 29–46.

[Hol16] Holmes, E. et al. “Applications of time-series analysis to mood fluctuations in bipolar disorder to promote treatment innovation: a case series”. *Translational Psychiatry* **6**.1 (2016), e720.

[HL17] Hong, M. & Luo, Z.-Q. “On the linear convergence of the alternating direction method of multipliers”. *Mathematical Programming* **162**.1-2 (2017), pp. 165–199.

[Hoo12] Hootman, J. M. et al. “A public health approach to addressing arthritis in older adults: the most common cause of disability”. *American journal of public health* **102**.3 (2012), pp. 426–433.

[How60] Howard, R. A. “Dynamic programming and markov processes.” (1960).

[Hua18] Huang, Q. et al. “Hidden markov models for monitoring circadian rhythmicity in telemetric activity data”. *Journal of The Royal Society Interface* **15**.139 (2018), p. 20170885.

[JP03] J Pineau G Gordon, S. T. “Point-based value iteration: An anytime algorithm for POMDPs”. *International Joint Conference on Artificial Intelligence* (2003).

[JP99] Jensen, J. L., Petersen, N. V., et al. “Asymptotic normality of the maximum likelihood estimator in state space models”. *The Annals of Statistics* **27**.2 (1999), pp. 514–535.

[Ji07] Ji, S. et al. “Point-based policy iteration”. *AAAI*. 2007, pp. 1243–1249.

[Jia19] Jiang, B. et al. “Entropy Learning for Dynamic Treatment Regimes”. *Statistica Sinica* (2019).

[JL15] Jiang, N. & Li, L. “Doubly robust off-policy value evaluation for reinforcement learning”. *arXiv preprint arXiv:1511.03722* (2015).

[Kae98] Kaelbling, L. P. et al. “Planning and acting in partially observable stochastic domains”. *Artificial intelligence* **101**.1-2 (1998), pp. 99–134.

[Kan18] Kanai, M. et al. “Effect of accelerometer-based feedback on physical activity in hospitalized patients with ischemic stroke: a randomized controlled trial”. *Clinical rehabilitation* (2018), p. 0269215518755841.

[Kos08] Kosorok, M. R. *Introduction to empirical processes and semiparametric inference.* Springer, 2008.

[Kup19] Kuppevelt, D. van et al. "Segmenting accelerometer data from daily life with unsupervised machine learning". *PloS one* **14**.1 (2019), e0208692.

[Kur86] Kurano, M. "Markov decision processes with a Borel measurable cost function?the average case". *Mathematics of operations research* **11**.2 (1986), pp. 309–320.

[Laa06] Laan, M. J. van der. "Causal effect models for intention to treat and realistic individualized treatment rules" (2006).

[LZ15] Laber, E. & Zhao, Y. "Tree-based methods for individualized treatment regimes". *Biometrika* **102**.3 (2015), pp. 501–514.

[LS17] Laber, E. B. & Staicu, A.-M. "Functional feature construction for individualized treatment regimes". *Journal of the American Statistical Association* just-accepted (2017).

[Lab14] Laber, E. B. et al. "Dynamic treatment regimes: Technical challenges and applications". *Electronic journal of statistics* **8**.1 (2014), p. 1225.

[LZ11] Langrock, R. & Zucchini, W. "Hidden Markov models with arbitrary state dwell-time distributions". *Computational Statistics & Data Analysis* **55**.1 (2011), pp. 715–724.

[Las08] Lascelles, B. D. X. et al. "Evaluation of a digitally integrated accelerometer-based activity monitor for the measurement of activity in cats". *Veterinary anaesthesia and analgesia* **35**.2 (2008), pp. 173–183.

[Las10] Lascelles, B. D. X. et al. "Cross-sectional study of the prevalence of radiographic degenerative joint disease in domesticated cats". *Veterinary surgery* **39**.5 (2010), pp. 535–544.

[Las17] Lascelles, B. D. X. et al. "Spontaneous painful disease in companion animals can facilitate the development of chronic pain therapies for humans". *Osteoarthritis and cartilage* (2017).

[LGM00] Le Gland, F. & Mevel, L. "Exponential forgetting and geometric ergodicity in hidden Markov models". *Mathematics of Control, Signals and Systems* **13**.1 (2000), pp. 63–93.

[Lee08] Lee, J.-G. et al. "Trajectory outlier detection: A partition-and-detect framework". *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on.* IEEE. 2008, pp. 140–149.

[LG16] Lee, J. A. & Gill, J. "Missing value imputation for physical activity data measured by accelerometer". *Statistical methods in medical research* (2016), p. 09622.

[LC11] Lee, Y.-S. & Cho, S.-B. "Activity recognition using hierarchical hidden markov models on a smartphone with 3D accelerometer". *International Conference on Hybrid Artificial Intelligence Systems*. Springer. 2011, pp. 460–467.

[LB17] Leos-Barajas, V. et al. "Analysis of animal accelerometer data using hidden Markov models". *Methods in Ecology and Evolution* **8**.2 (2017), pp. 161–173.

[Ler92] Leroux, B. G. "Maximum-likelihood estimation for hidden Markov models". *Stochastic processes and their applications* **40**.1 (1992), pp. 127–143.

[Lev86] Levinson, S. E. "Continuously variable duration hidden Markov models for automatic speech recognition". *Computer Speech & Language* **1**.1 (1986), pp. 29–45.

[Lia19] Liao, P. et al. "Off-Policy Estimation of Long-Term Average Outcomes with Applications to Mobile Health". *arXiv preprint arXiv:1912.13088* (2019).

[Lit94] Littman, M. L. "The witness algorithm: Solving partially observable Markov decision processes". *Brown University, Providence, RI* (1994).

[Liu19] Liu, N. et al. "Learning the Dynamic Treatment Regimes from Medical Registry Data through Deep Q-network". *Scientific reports* **9**.1 (2019), p. 1495.

[Liu15] Liu, Y.-Y. et al. "Efficient learning of continuous-time hidden markov models for disease progression". *Advances in neural information processing systems*. 2015, pp. 3600–3608.

[Lu13] Lu, W. et al. "Variable selection for optimal treatment decision". *Statistical methods in medical research* **22**.5 (2013), pp. 493–504.

[Luc19] Luckett, D. J. et al. "Estimating dynamic treatment regimes in mobile health using V-learning". *Journal of the American Statistical Association* (2019), pp. 1–34.

[LL16] Luedtke, A. R. & Laan, M. J. van der. "Super-learning of an optimal dynamic treatment rule". *The international journal of biostatistics* **12**.1 (2016), pp. 305–332.

[MS10] Mannini, A. & Sabatini, A. M. "Machine learning methods for classifying human physical activity from on-body accelerometers". *Sensors* **10**.2 (2010), pp. 1154–1175.

[Mar07] Marshall, A. et al. "Self management of chronic disease using mobile devices and Bluetooth monitors". *Proceedings of the ICST 2nd international confer-*

ence on Body area networks. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). 2007, p. 22.

[McC16] McClintock, M. K. et al. “Empirical redefinition of comprehensive health and well-being in the older adults of the United States”. *Proceedings of the National Academy of Sciences* **113**.22 (2016), E3071–E3080.

[MQ14] McKeague, I. W. & Qian, M. “Estimation of treatment policies based on functional predictors”. *Statistica Sinica* **24**.3 (2014), p. 1461.

[Met08] Metzger, J. S. et al. “Patterns of objectively measured physical activity in the United States.” *Medicine and science in sports and exercise* **40**.4 (2008), pp. 630–638.

[MT12] Meyn, S. P. & Tweedie, R. L. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[Mog09] Mogil, J. S. “Animal models of pain: progress and challenges”. *Nature Reviews Neuroscience* **10**.4 (2009), pp. 283–294.

[Mon82] Monahan, G. E. “State of the art—a survey of partially observable Markov decision processes: theory, models, and algorithms”. *Management Science* **28**.1 (1982), pp. 1–16.

[Moo07] Moodie, E. E. et al. “Demystifying optimal dynamic treatment regimes”. *Biometrics* **63**.2 (2007), pp. 447–455.

[Moo14a] Moodie, E. E. et al. “Q-learning: Flexible learning about useful utilities”. *Statistics in Biosciences* **6**.2 (2014), pp. 223–243.

[Moo12] Moore, P. J. et al. “Forecasting depression in bipolar disorder”. *IEEE Transactions on Biomedical Engineering* **59**.10 (2012), pp. 2801–2807.

[Moo14b] Moore, P. J. et al. “Mood dynamics in bipolar disorder”. *International journal of bipolar disorders* **2**.1 (2014), p. 11.

[Mor06] Morris, J. S. et al. “Using wavelet-based functional mixed models to characterize population heterogeneity in accelerometer profiles: a case study”. *Journal of the American Statistical Association* **101**.476 (2006), pp. 1352–1364.

[Mur03] Murphy, S. A. “Optimal dynamic treatment regimes”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**.2 (2003), pp. 331–355.

[Mur05] Murphy, S. A. “A generalization error for Q-learning”. *Journal of Machine Learning Research* **6**.Jul (2005), pp. 1073–1097.

[Mur01] Murphy, S. A. et al. “Marginal mean models for dynamic regimes”. *Journal of the American Statistical Association* **96**.456 (2001), pp. 1410–1423.

[Mur16] Murphy, S. A. et al. "A batch, off-policy, actor-critic algorithm for optimizing the average reward". *arXiv preprint arXiv:1607.05047* (2016).

[Nap10] Napolitano, M. A. et al. "Accelerometer use in a physical activity intervention trial". *Contemporary clinical trials* **31**.6 (2010), pp. 514–523.

[NN07] Natarajan, P. & Nevatia, R. "Coupled hidden semi markov models for activity recognition". *Motion and Video Computing, 2007. WMVC'07. IEEE Workshop on*. IEEE. 2007, pp. 10–10.

[Nic11] Nickel, C. et al. "Using hidden markov models for accelerometer-based biometric gait recognition". *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*. IEEE. 2011, pp. 58–63.

[Nod12] Nodelman, U. et al. "Expectation maximization and complex duration distributions for continuous time Bayesian networks". *arXiv preprint arXiv:1207.1402* (2012).

[OR12] Olteanu, M. & Ridgway, J. "Hidden Markov models for time series of counts with excess zeros". *European Symposium on Artificial Neural Networks*. 2012, pp. 133–138.

[Par18] Park, J. et al. "Various Types of Arthritis in the United States: Prevalence and Age-Related Trends From 1999 to 2014". *American journal of public health* 0 (2018), e1–e3.

[Pat15] Patel, R. et al. "Do antidepressants increase the risk of mania and bipolar disorder in people with depression? A retrospective electronic case register cohort study". *BMJ open* **5**.12 (2015), e008341.

[PP13] Pötscher, B. M. & Prucha, I. R. *Dynamic nonlinear econometric models: Asymptotic theory*. Springer Science & Business Media, 2013.

[Pow07] Powell, W. B. *Approximate Dynamic Programming: Solving the curses of dimensionality*. Vol. 703. John Wiley & Sons, 2007.

[Put14] Puterman, M. L. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[Pyk61a] Pyke, R. "Markov renewal processes: definitions and preliminary properties". *The Annals of Mathematical Statistics* (1961), pp. 1231–1242.

[Pyk61b] Pyke, R. "Markov renewal processes with finitely many states". *The Annals of Mathematical Statistics* (1961), pp. 1243–1259.

[Rab89] Rabiner, L. R. "A tutorial on hidden Markov models and selected applications in speech recognition". *Proceedings of the IEEE* **77**.2 (1989), pp. 257–286.

[Rob10] Robertson, W. et al. "Utility of accelerometers to measure physical activity in children attending an obesity treatment intervention". *Journal of obesity* **2011** (2010).

[Rob08] Robins, J. et al. "Estimation and extrapolation of optimal treatment and testing strategies". *Statistics in medicine* **27**.23 (2008), pp. 4678–4721.

[Rob97] Robins, J. M. "Causal inference from complex longitudinal data". *Latent variable modeling and applications to causality*. Ed. by Berkane, M. New York: Springer Science & Business Media, 1997. Chap. 4, pp. 69–117.

[Rob99] Robins, J. M. "Testing and estimation of direct effects by reparameterizing directed acyclic graphs with structural nested models". *Computation, causation, and discovery* (1999), pp. 349–405.

[Rob04] Robins, J. M. "Optimal structural nested models for optimal sequential decisions". *Proceedings of the second seattle Symposium in Biostatistics*. Springer. 2004, pp. 189–326.

[RC14] Ronao, C. A. & Cho, S.-B. "Human activity recognition using smartphone sensors with two-stage continuous hidden Markov models". *Natural Computation (ICNC), 2014 10th International Conference on*. IEEE. 2014, pp. 681–686.

[RL12] Rubin, D. B. & Laan, M. J. van der. "Statistical issues and limitations in personalized medicine research with clinical trials". *The international journal of biostatistics* **8**.1 (2012).

[Rub04] Rubin, D. B. *Multiple imputation for nonresponse in surveys*. Vol. 81. John Wiley & Sons, 2004.

[RM85] Russell, M. & Moore, R. "Explicit modelling of state occupancy in hidden Markov models for automatic speech recognition". *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'85.* Vol. 10. IEEE. 1985, pp. 5–8.

[Sac03] Sachs, G. S. et al. "Rationale, design, and methods of the systematic treatment enhancement program for bipolar disorder (STEP-BD)". *Biological psychiatry* **53**.11 (2003), pp. 1028–1042.

[Sch15] Schmid, D. et al. "Associations of objectively assessed physical activity and sedentary time with all-cause mortality in US adults: the NHANES study". *PLoS One* **10**.3 (2015), e0119591.

[Sch14] Schulte, P. J. et al. "Q-and A-learning methods for estimating optimal dynamic treatment regimes". *Statistical science: a review journal of the Institute of Mathematical Statistics* **29**.4 (2014), p. 640.

[Sch78] Schwarz, G. et al. “Estimating the dimension of a model”. *The annals of statistics* **6**.2 (1978), pp. 461–464.

[Sco05] Scott, S. L. et al. “Hidden Markov models for longitudinal comparisons”. *Journal of the American Statistical Association* **100**.470 (2005), pp. 359–369.

[Shi18] Shi, C. et al. “HIGH-DIMENSIONAL A-LEARNING FOR OPTIMAL DYNAMIC TREATMENT REGIMES.” *Annals of statistics* **46**.3 (2018), pp. 925–957.

[Shi19] Shi, C. et al. “A sparse random projection-based test for overall qualitative treatment effects”. *Journal of the American Statistical Association* just-accepted (2019), pp. 1–41.

[Shi14] Shi, W. et al. “On the Linear Convergence of the ADMM in Decentralized Consensus Optimization.” *IEEE Trans. Signal Processing* **62**.7 (2014), pp. 1750–1761.

[Shi10] Shirley, K. E. et al. “Hidden Markov models for alcoholism treatment trial data”. *The Annals of Applied Statistics* (2010), pp. 366–395.

[SV10] Silver, D. & Veness, J. “Monte-Carlo planning in large POMDPs”. *Advances in neural information processing systems*. 2010, pp. 2164–2172.

[Sli11] Slingerland, L. et al. “Cross-sectional study of the prevalence and clinical features of osteoarthritis in 100 cats”. *The Veterinary Journal* **187**.3 (2011), pp. 304–309.

[SS12] Smith, T. & Simmons, R. “Heuristic search value iteration for POMDPs”. *arXiv preprint arXiv:1207.4166* (2012).

[So98] So, M. E. P. et al. “A stochastic volatility model with Markov switching”. *Journal of Business & Economic Statistics* **16**.2 (1998), pp. 244–253.

[Son78] Sondik, E. J. “The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs”. *Operations research* **26**.2 (1978), pp. 282–304.

[Son71] Sondik, E. J. *The optimal control of partially observable Markov processes.* Tech. rep. Stanford Univ Calif Stanford Electronics Labs, 1971.

[Son15] Song, R. et al. “Penalized q-learning for dynamic treatment regimens”. *Statistica Sinica* **25**.3 (2015), p. 901.

[SD97] Stukel, T. A. & Demidenko, E. “Two-stage method of estimation for general linear growth curve models”. *Biometrics* (1997), pp. 720–728.

[SB18] Sutton, R. S., Barto, A. G., et al. *Reinforcement learning: An introduction*. MIT press, 2018.

[Tao18] Tao, Y. et al. "Tree-based reinforcement learning for estimating optimal dynamic treatment regimes". *The annals of applied statistics* **12**.3 (2018), p. 1914.

[Tay15] Taylor, J. M. et al. "Reader reaction to âĂIJA robust method for estimating optimal treatment regimesâĂİ by Zhang et al.(2012)". *Biometrics* **71**.1 (2015), pp. 267–273.

[TM17] Tewari, A. & Murphy, S. A. "From ads to interventions: Contextual bandits in mobile health". *Mobile Health*. Springer, 2017, pp. 495–517.

[Tho10] Thomas, O. et al. "Wearable sensor activity analysis using semi-Markov models with a grammar". *Pervasive and Mobile Computing* **6**.3 (2010), pp. 342–350.

[TB16] Thomas, P. & Brunskill, E. "Data-efficient off-policy policy evaluation for reinforcement learning". *International Conference on Machine Learning*. 2016, pp. 2139–2148.

[Tia14] Tian, L. et al. "A simple method for estimating interactions between a treatment and a large number of covariates". *Journal of the American Statistical Association* **109**.508 (2014), pp. 1517–1532.

[TL11] Trevezas, S. & Limnios, N. "Exact MLE and asymptotic properties for nonparametric semi-Markov models". *Journal of Nonparametric Statistics* **23**.3 (2011), pp. 719–739.

[Tro08] Troiano, R. P. et al. "Physical activity in the United States measured by accelerometer". *Medicine and science in sports and exercise* **40**.1 (2008), p. 181.

[Tsi19] Tsiatis, A. et al. *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine*. CRC press, 2019.

[Tur08] Turner, R. "Direct maximization of the likelihood of a hidden Markov model". *Computational Statistics & Data Analysis* **52**.9 (2008), pp. 4147–4160.

[VDV91] Van Der Vaart, A. "On differentiable functionals". *The Annals of Statistics* (1991), pp. 178–204.

[VDVW96] Van Der Vaart, A. W. & Wellner, J. A. "Weak convergence". *Weak convergence and empirical processes*. Springer, 1996, pp. 16–28.

[VM09] Verbeke, G. & Molenberghs, G. *Linear mixed models for longitudinal data*. Springer Science & Business Media, 2009.

[Vie08] Vierck, C. et al. "Clinical and pre-clinical pain assessment: Are we measuring the same thing?" *Pain* **135**.1-2 (2008), pp. 7–10.

[Vit67] Viterbi, A. “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm”. *IEEE transactions on Information Theory* **13**.2 (1967), pp. 260–269.

[WA18] Wager, S. & Athey, S. “Estimation and inference of heterogeneous treatment effects using random forests”. *Journal of the American Statistical Association* **113**.523 (2018), pp. 1228–1242.

[Wal19] Wallace, M. P. et al. “Model selection for G-estimation of Dynamic Treatment Regimes”. *Biometrics* (2019).

[Wan14] Wang, X. et al. “Unsupervised learning of disease progression models”. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2014, pp. 85–94.

[WVO12] Wiering, M. & Van Otterlo, M. “Reinforcement learning”. *Adaptation, learning, and optimization* **12** (2012), p. 3.

[Wit14] Witowski, V. et al. “Using Hidden Markov Models to Improve Quantifying Physical Activity in Accelerometer Data–A Simulation Study”. *PloS one* **9**.12 (2014), e114089.

[Wu15] Wu, F. et al. “Who will benefit from antidepressants in the acute treatment of bipolar depression? A reanalysis of the STEP-BD study by Sachs et al. 2007, using Q-learning”. *International journal of bipolar disorders* **3**.1 (2015), p. 7.

[Xia14] Xiao, L. et al. “Quantifying the lifetime circadian rhythm of physical activity: a covariate-dependent functional approach”. *Biostatistics* **16**.2 (2014), pp. 352–367.

[Xu16] Xu, Y. et al. “Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times”. *Journal of the American Statistical Association* **111**.515 (2016), pp. 921–950.

[Yam75] Yamada, K. “Duality theorem in Markovian decision problems”. *Journal of mathematical analysis and applications* **50**.3 (1975), pp. 579–595.

[Yu15] Yu, S.-Z. *Hidden Semi-Markov Models: Theory, Algorithms and Applications*. Morgan Kaufmann, 2015.

[YK03] Yu, S.-Z. & Kobayashi, H. “A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking”. *Signal Processing* **83**.2 (2003), pp. 235–250.

[Zha12a] Zhang, B. et al. “Estimating optimal treatment regimes from a classification perspective”. *Stat* **1**.1 (2012), pp. 103–114.

[ZL96] Zhang, N. L. & Liu, W. *Planning in stochastic domains: Problem characteristics and approximation*. Tech. rep. Technical Report HKUST-CS96-31, Hong Kong University of Science and Technology, 1996.

[Zha18] Zhang, Y. et al. "Interpretable dynamic treatment regimes". *Journal of the American Statistical Association* **113**.524 (2018), pp. 1541–1549.

[Zha15] Zhao, Y.-Q. et al. "New statistical learning methods for estimating optimal dynamic treatment regimes". *Journal of the American Statistical Association* **110**.510 (2015), pp. 583–598.

[Zha12b] Zhao, Y. et al. "Estimating individual treatment rules using outcome weighted learning". *Journal of the American Statistical Association* **107**.499 (2012), p. 1106.

[Zha09] Zhao, Y. et al. "Reinforcement learning design for cancer clinical trials". *Statistics in medicine* **28**.26 (2009), pp. 3294–3315.

[ZH01] Zhou, R. & Hansen, E. A. "An improved grid-based approximation algorithm for POMDPs". *IJCAI*. 2001, pp. 707–716.

[ZS09] Zhu, C. & Sheng, W. "Human daily activity recognition in robot-assisted living using multi-sensor fusion". *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE. 2009, pp. 2154–2159.

APPENDICES

APPENDIX

A

SUPPLEMENTARY MATERIALS FOR CHAPTER 2

A.1 Technical details and proofs

A.1.1 Maximum likelihood estimation for ZIPHSMM

The EM algorithm [Bau70] is the most commonly used approach to fit HMMs. However, the EM algorithm is known to suffer from slow convergence especially when the likelihood function is flat [Tur08]. Further complicating computation is that we would have to embed an additional layer of EM algorithm for the zero-inflated Poisson regressions where the M-step would be the Newton-Raphson update, which is rather cumbersome. Instead, we perform parameter estimation by directly maximizing the log-likelihood from the forward algorithm using a Quasi-Newton method. This method is straightforward and enjoys the local superlinear convergence. To avoid local maxima, three different sets of initial values are used for the ZIPHSMM parameters.

To carry out the fitting procedure, we reformulate ZIPHSMM as ZIPHMM with an expanded state space [LZ11]. This is done by redefining the latent states as the tuple of latent states and state durations. The prior probabilities are reparameterized using generalized logits with state 1 as the reference; the latent state durations are parameterized by accelerated failure time models; the transition probabilities are reparameterized as multinomial logistic regressions; the state-dependent zero proportion and Poisson means are parameterized by generalized linear models with logit and log links. These specifications are consistent with the description of the ZIPHSMM in the paper.

A.1.2 Identifiability of the zero-inflated hidden Markov model

Proof. According to Proposition 2.1 in [Gas13], if the state-dependent probability measures are linearly independent and the weight matrix has full rank, then the parameters in the hidden Markov model are identifiable up to label switching of the hidden states as soon as the transition matrix has full rank. In our case, define the M state-dependent probability measures $\mu_1, ..., \mu_M$ by

$$\mu_j = \sum_{l=1}^{k} \psi_{jl} \phi_l \text{ for } j = 1, ..., M,$$

where ϕ_1 is a point mass at zero, ϕ_l (l=2,...,M+1) are the M different Poisson distributions, and $\psi = [\psi_{jl}]$ is the weight matrix. Denote the structural zero proportion for the zero-inflated Poisson distribution in state 1 by p. Then the $M \times (M+1)$ weight matrix ψ is

$$\psi = \begin{bmatrix} p & 1-p & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & \dots & \dots & \dots & \dots & 1 \end{bmatrix},$$

where each row represents a latent state; column 1 represents the probability measure of the structural zeros; column 2 - column M+1 represent the M distinct Poisson distributions. It can be seen that ψ has rank M.

A.1.3 Proof of Lemma 2.3.1

Define the conditional log likelihood function for the subject-specific hidden semi-Markov model given initial state S_i^0 as

$$l_i^T(\theta_i, S_i^0) \triangleq \log P(Y_i^1, ..., Y_i^T | S_i^0, \theta_i) = \sum_{t=1}^{T} \log P(Y_i^t | Y_i^0, ..., Y_i^{t-1}, S_i^0, \theta_i)$$

for subject $i = 1, ..., n$. Under (A1), this is equivalent to the conditional log likelihood for a correspondingly reparameterized hidden Markov model with an expanded state space.

Under (A0) - (A4), the uniform convergence of the normalized log likelihood is proved in Proposition 2 in Douc et al. [Dou04], i.e.,

$$\sup_{\theta_i} \sup_{S_i^0} |T^{-1} l_i^T(\theta_i, S_i^0) - l_i(\theta_i)| \to 0 \; a.s. \text{ as } T \to \infty \text{ for } i = 1, ..., n,$$

where $l_i(\theta_i)$ is a deterministic asymptotic function such that

$$l_i(\theta_i) \triangleq \lim_{k\to\infty} E_{\theta_i^\star}\left[\log P(Y_i^{\,t} | Y_i^{-k}, ..., Y_i^{\,t-1}, S_i^{-m}, \theta_i)\right].$$

To show that $l_i(\theta_i)$ is maximized at $\theta_i = \theta_i^\star$ for $i = 1, ..., n$,

$$\begin{aligned}
& l_i(\theta_i) - l_i(\theta_i^\star) \\
=& \lim_{k\to\infty} E_{\theta_i^\star}\left[\log \frac{P(Y_i^{\,t} | Y_i^{-k}, ..., Y_i^{\,t-1}, S_i^{-m}, \theta_i)}{P(Y_i^{\,t} | Y_i^{-k}, ..., Y_i^{\,t-1}, S_i^{-m}, \theta_i^\star)}\right] \\
=& \lim_{k\to\infty} E_{\theta_i^\star}\left\{ E_{\theta_i^\star}\left[\log \frac{P(Y_i^{\,t} | Y_i^{-k}, ..., Y_i^{\,t-1}, S_i^{-m}, \theta_i)}{P(Y_i^{\,t} | Y_i^{-k}, ..., Y_i^{\,t-1}, S_i^{-m}, \theta_i^\star)} \Big| Y_i^{-k}, ..., Y_i^{\,t-1}, S_i^{-m}, \theta_i^\star\right]\right\},
\end{aligned}$$

where the inner conditional expectation is the Kullback-Leibler divergence and is maximized at $\theta_i = \theta_i^\star$.

By the identifiability assumption (A5), $\theta_i^\star$ is the unique maximizer of $l_i(\theta_i)$. Therefore, it follows that

$$\arg\max_{\theta_i} T^{-1} l_i^T(\theta_i, S_i^0) \to \arg\max_{\theta_i} l_i(\theta_i) \text{ with probability 1 as } T \to \infty.$$

Because $\arg\max_{\theta_i} T^{-1} l_i^T(\theta_i, S_i^0)$ is the MLE $\hat{\theta}_{i,T}$, we have

$$\hat{\theta}_{i,T} \to \theta_i^\star \text{ with probability 1 for all } i = 1, ..., n.$$

A.1.4 Proof of Theorem 2.3.1

It suffices to proof the theorem for $\widehat{\mathbf{\Omega}}_{m,n}$, which can then be naturally extended to $\widehat{\mathbf{\Gamma}}_{m,n}$ and $\widehat{\mathbf{\Lambda}}_{m,\ell,n}$ using the exact same approach, where $m, \ell = 1, ..., M$. By Assumptions (A0) - (A5), we have Lemma 1, which implies that the MLEs for subject-specific treatment effects $\widehat{\boldsymbol{b}}_{1,m,T}$ from the first-stage fitting are strongly consistent; this implies that

$$\mathrm{Var}[\mathbf{vec}(\widehat{\boldsymbol{b}}_{1,m,T})_i] \longrightarrow 0 \;\text{ for all } i = 1, ..., qn,\; m = 1, ..., M \text{ as } T \longrightarrow \infty.$$

Recall that the linear regression model we fit in the second stage is of the form

$$\mathbf{vec}(\widehat{\boldsymbol{b}}_{1,m,T})_i = \tilde{\boldsymbol{B}}_i'\mathbf{vec}(\Omega_m) + v_{m,i} \text{ for } m=1,\ldots,M,\ i=1,\ldots,qn$$

where $\tilde{\boldsymbol{B}} \triangleq \boldsymbol{I} \otimes \tilde{\boldsymbol{W}} \in \mathbb{R}^{qn\times q(d+1)}$ the i^{th} row of which is $\tilde{\boldsymbol{B}}_i$; $v_{m,i} = u_{m,i} + e_{m,i}$; $e_{m,i}$ and is independent across all indices with mean 0 and variance σ^2; $u_{m,i}$ and furthermore is independent across all indices with mean 0 and variance $\text{Var}[\mathbf{vec}(\widehat{b}_{1,m,T})_i]$. In the corresponding matrix notation,

$$\mathbf{vec}(\widehat{\boldsymbol{b}}_{1,m,T}) = \tilde{\boldsymbol{B}}\,\mathbf{vec}(\Omega_m) + \boldsymbol{v}_m \text{ for } m=1,\ldots,M.$$

Consider the true linear model for $\boldsymbol{b}_{1,m}$ as

$$\mathbf{vec}(\boldsymbol{b}_{1,m}) = \tilde{\boldsymbol{B}}\,\mathbf{vec}(\Omega_m^*) + \boldsymbol{e}_m \text{ for } m=1,\ldots,M,$$

where $\mathbf{vec}(\widehat{\boldsymbol{b}}_{1,m,T}) = \mathbf{vec}(\boldsymbol{b}_{1,m}) + \boldsymbol{u}_m$ such that $\boldsymbol{u}_{m,k}$ mimics the measurement error process depending only on the first-stage fit. We want to show $\mathbf{vec}(\widehat{\Omega}_{m,n}) \xrightarrow{p} \mathbf{vec}(\Omega_m^*)$ as $n \longrightarrow \infty$. Because the first-stage subject covariates and environmental factors are not correlated with the second-stage baseline characteristics, we have $\boldsymbol{u}_m \perp \boldsymbol{e}_m$, $\boldsymbol{u}_m \perp \tilde{\boldsymbol{B}}$, and $\boldsymbol{e}_m \perp \tilde{\boldsymbol{B}}$, so that $E(\tilde{\boldsymbol{B}}^T\boldsymbol{u}_m) = 0$, $E(\tilde{\boldsymbol{B}}^T\boldsymbol{e}_m) = 0$ for $m=1,\ldots,M$.

We also assume that $\lim_{n\to\infty} \frac{1}{n}\tilde{\boldsymbol{W}}^T\tilde{\boldsymbol{W}} = \mathbf{H}$ where $\mathbf{H}$ is a positive definite matrix. Since $E\left(\frac{\tilde{\boldsymbol{B}}^T\boldsymbol{v}_m}{n}\right) = E\left(\frac{\tilde{\boldsymbol{B}}^T(\boldsymbol{u}_m+\boldsymbol{e}_m)}{n}\right) = 0$ and $E(v_{m,i}^2) = E(u_{m,i}^2) + E(e_{m,i}^2) = \text{Var}[\mathbf{vec}(\widehat{b}_{1,m,T})_i] + \sigma^2 \xrightarrow{p} \sigma^2$, then $Var(v_{m,i}) \le g\sigma^2$ for all $i=1,\ldots,nq$ when T is large, where $g>1$ is some fixed positive constant. Then

$$Var\left(\frac{\tilde{\boldsymbol{B}}^T\boldsymbol{v}_m}{n}\right) = \frac{\tilde{\boldsymbol{B}}^T Var(\boldsymbol{v}_m)\tilde{\boldsymbol{B}}}{n^2} \le \frac{g\sigma^2}{n}\frac{\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}}{n} = \frac{g\sigma^2}{n}\left[\mathbf{I}\otimes\frac{1}{n}\tilde{\boldsymbol{W}}^T\tilde{\boldsymbol{W}}\right] \to 0 \text{ as } n\to\infty,$$

which implies that $\frac{\tilde{\boldsymbol{B}}^T\boldsymbol{v}_m}{n} \xrightarrow{p} 0$. Therefore, we can show

$$\begin{aligned}
\mathbf{vec}(\widehat{\Omega}_{m,n}) &= \left(\frac{\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}}{n}\right)^{-1}\frac{\tilde{\boldsymbol{B}}^T\mathbf{vec}(\widehat{\boldsymbol{b}}_{1,m,T})}{n}\\
&= \left(\frac{\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}}{n}\right)^{-1}\frac{\tilde{\boldsymbol{B}}^T(\tilde{\boldsymbol{B}}\mathbf{vec}(\Omega_m^*) + \boldsymbol{v}_m)}{n}\\
&= \mathbf{vec}(\Omega_m^*) + \left(\frac{\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}}{n}\right)^{-1}\frac{\tilde{\boldsymbol{B}}^T\boldsymbol{v}_m}{n}\\
&\xrightarrow{p} \mathbf{vec}(\Omega_m^*)
\end{aligned}$$

for $m = 1, ..., M$ as $n \longrightarrow \infty$. Thus, $\mathbf{vec}(\widehat{\boldsymbol{\Omega}}_{m,n}) \xrightarrow{p} \mathbf{vec}(\boldsymbol{\Omega}^*_m)$ for $m = 1, ..., M$ as $n \longrightarrow \infty$.

By Lindeberg-Feller Central Limit Theorem, $\sqrt{n}(\mathbf{vec}(\widehat{\boldsymbol{\Omega}}_{m,n}) - \mathbf{vec}(\boldsymbol{\Omega}^*_m))$ converges in distribution to a multivariate Gaussian distribution with mean $\mathbf{0}$ and covariance

$$(\mathbf{I} \otimes \mathbf{H})^{-1} \left[\frac{1}{n} \tilde{\boldsymbol{B}}^T \mathbf{diag}\{E(v^2_{m,i})\} \tilde{\boldsymbol{B}} \right] (\mathbf{I} \otimes \mathbf{H})^{-1},$$

where $\mathbf{diag}\{E(v^2_{m,i})\}$ is a diagonal matrix with elements $\{E(v^2_{m,i})\}$.

By the consistency of the sandwich variance estimator,

$$\left(\left[\frac{\tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}}}{n} \right]^{-1} \left[\frac{1}{n} \tilde{\boldsymbol{B}}^T \mathbf{diag} \left\{ [\mathbf{vec}(\widehat{\boldsymbol{b}}_{1,m,T})_i - \tilde{\boldsymbol{B}}_i^T \mathbf{vec}(\boldsymbol{\Omega}^*_m)]^2 \right\} \tilde{\boldsymbol{B}} \right] \left[\frac{\tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}}}{n} \right]^{-1} \right)$$
$$\rightarrow (\mathbf{I} \otimes \mathbf{H})^{-1} \left[\frac{1}{n} \tilde{\boldsymbol{B}}^T \mathbf{diag}\{E(v^2_{m,i})\} \tilde{\boldsymbol{B}} \right] (\mathbf{I} \otimes \mathbf{H})^{-1}$$

Thus, by Slutsky's Theorem, the following quantity converges in distribution to a multivariate Gaussian distribution with mean zero and identity covariance,

$$\left(\left[\frac{\tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}}}{n} \right]^{-1} \left[\frac{1}{n} \tilde{\boldsymbol{B}}^T \mathbf{diag} \left\{ [\mathbf{vec}(\widehat{\boldsymbol{b}}_{1,m,T})_i - \tilde{\boldsymbol{B}}_i^T \mathbf{vec}(\widehat{\boldsymbol{\Omega}}_{m,n})]^2 \right\} \tilde{\boldsymbol{B}} \right] \left[\frac{\tilde{\boldsymbol{B}}^T \tilde{\boldsymbol{B}}}{n} \right]^{-1} \right)^{-1/2} \times$$
$$\sqrt{n} \left\{ \mathbf{vec}(\widehat{\boldsymbol{\Omega}}_{m,n}) - \mathbf{vec}(\boldsymbol{\Omega}^*_m) \right\}.$$

A.2 Runtime evaluation of the ZIPHSMM

Figure A.1 shows the computational cost associated with the proposed two-stage estimator under the generative model in Scenario I as described in the simulation experiments. We simulated 10, 20, and 40 independent series of minute-by-minute activity count data for 8 weeks, 16 weeks, and 32 weeks. It can be seen that the algorithm scales linearly with the length of the series and the number of subjects. However, there is a universal caveat in the scalability for the forward-backward algorithm in hidden semi-Markov models, which is that the computational cost scales quadratically with the number of hidden states (M) and the maximum latent state duration (D). Based on the results in [Rab89], the computational cost of a single hidden semi-Markov model would be $O(M^2 D^2 T)$, where M is the number of latent states, D is the maximum latent state duration, and T is the length of the observed series. Because our model is fit seperately for each of the n subjects in the first stage, the total computational cost is of order $O(nM^2 D^2 T)$.

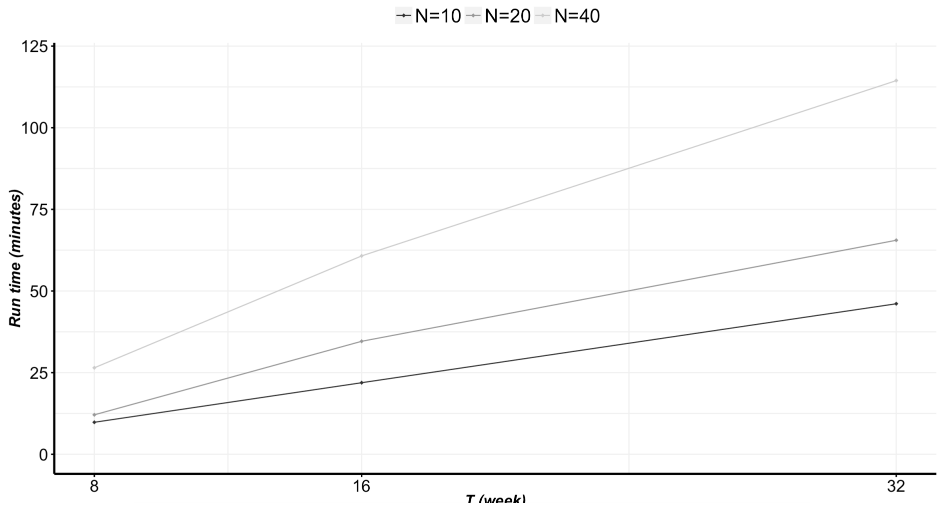

Figure A.1 Runtimes of fitting the subject-specific ZIPHSMM for 10, 20, and 40 subjects with the length of series equal to 8, 16, and 32 weeks. The algorithm scales linearly with the length of the series and the number of subjects.

APPENDIX

B

SUPPLEMENTARY MATERIALS FOR CHAPTER 3

B.1 Technical details and proofs

B.1.1 An Example of Constraint Matrices in ADMM

Suppose: (i) the intercepts in the generalized linear models for state-dependent parameters are subject-specific; (ii) the initial state probabilities and the transition rate parameters are group-specific; (iii) the slope parameters in the generalized linear models for state-dependent parameters to be common across all subjects. The problem of minimizing the negative log likelihood with respect to linear constraints is specified as follows,

$$
\begin{aligned}
&\min_{\boldsymbol{\theta},\mathbf{z}} f(\boldsymbol{\theta}) \\
&s.t.\ \mathbf{A}_i\boldsymbol{\theta}_i = \mathbf{B}_i\mathbf{z},\ i = 1,\dots,n.
\end{aligned}
$$

Here $\mathbf{z}$ represents the set of all group-specific and common parameters in $\boldsymbol{\theta}$ such that

$$
\mathbf{z} \triangleq [\mathbf{a}^{1'},\dots,\mathbf{a}^{K'},\mathbf{c}^{1'},\dots,\mathbf{c}^{K'},\mathbf{b}_1'] \in \mathbb{R}^{(M+1)(KM-K+q)},
$$

where the superscript stands for the group index. The constraint matrices $\mathbf{A}_i \in \mathbb{R}^{(M+1)(M+q-1)\times(M+1)(M+q)}$ and $\mathbf{B}_i \in \mathbb{R}^{(M+1)(M+q-1)\times(M+1)(KM-K+q)}$ ensure that the group-specific parameters and common parameters share the same values within group and across all subjects respectively. For example, if subject i belongs to group 1, then

$$
\begin{aligned}
\boldsymbol{\theta}_i &= [a_{i,1},\ldots,a_{i,M-1},c_{i,1,2},\ldots,c_{i,M,M-1},\ldots,b_{i,0,0},\ldots,b_{i,M,0},\mathbf{b}'_{i,0,1},\ldots,\mathbf{b}'_{i,M,1}],\\
\mathbf{z} &= [a_1^1,a_2^1\ldots,a_{M-1}^1,\ldots,a_1^K,\ldots,a_{M-1}^K,\ldots,a_{M-1}^K,c_{1,2}^1,\ldots,c_{M,M-1}^1,\ldots,c_{1,2}^K,\ldots,c_{M,M-1}^K,\mathbf{b}'_{0,1},\ldots,\mathbf{b}'_{M,1}],\\
\mathbf{A}_i &= \begin{bmatrix} \mathbf{I}_{(M+1)(M-1)} & \mathbf{0}_{(M+1)(M-1)\times(M+1)} & \mathbf{0}_{(M+1)(M-1)\times q(M+1)} \\ \mathbf{0}_{q(M+1)\times(M+1)(M-1)} & \mathbf{0}_{q(M+1)\times(M+1)} & \mathbf{I}_{q(M+1)} \end{bmatrix},\\
\mathbf{B}_i &= \begin{bmatrix} \mathbf{I}_{M-1} & \mathbf{0}_{(M-1)\times(K-1)(M-1)} & \mathbf{0}_{(M-1)\times M(M-1)} & \mathbf{0}_{(M-1)\times(K-1)M(M-1)} & \mathbf{0}_{(M-1)\times q(M+1)} \\ \mathbf{0}_{M(M-1)\times(M-1)} & \mathbf{0}_{M(M-1)\times(K-1)(M-1)} & \mathbf{I}_{M(M-1)} & \mathbf{0}_{M(M-1)\times(K-1)M(M-1)} & \mathbf{0}_{(M-1)\times q(M+1)} \\ \mathbf{0}_{q(M+1)\times(M-1)} & \mathbf{0}_{q(M+1)\times(K-1)(M-1)} & \mathbf{0}_{q(M+1)\times M(M-1)} & \mathbf{0}_{q(M+1)\times(K-1)M(M-1)} & \mathbf{I}_{q(M+1)} \end{bmatrix}
\end{aligned}
$$

B.1.2 Gradient computation for the CTHMM

The computation of the gradients in HMM is based on the Fisher's identity [Cap05]. Here we are omitting the subscript i in the notation because the derivation is done in the same fashion for all subjects $i = 1,\ldots,n$. Denote $\boldsymbol{\theta}$ to be the HMM working parameter for a generic subject. Denote the observed data log likelihood function by $l(\mathbf{y}|\boldsymbol{\theta})$, i.e., the marginal density of $\mathscr{Y} \triangleq Y(T_1),\ldots,Y(T_K)$. Denote by $u(\mathbf{s},\mathbf{y}|\boldsymbol{\theta})$ the complete data likelihood, i.e., the joint density of $Y(T_1),\ldots,Y(T_K)$ and $S(T_1),\ldots,S(T_K)$. Also, let $p(\mathbf{s}|\boldsymbol{\theta},\mathbf{y})$ be the conditional density of $S(T_1),\ldots,S_i(T_K)$ given $Y(T_1),\ldots,Y(T_K)$. Similar to the EM algorithm, define

$$Z(\boldsymbol{\theta},\boldsymbol{\theta}^*) = E_{\boldsymbol{\theta}^*}[\log u(\mathbf{s},\mathbf{y}|\boldsymbol{\theta})|\mathscr{Y}] = \int \log u(\mathbf{s},\mathbf{y}|\boldsymbol{\theta})p(\mathbf{s}|\mathbf{y},\boldsymbol{\theta}^*)d\mathbf{s}, \tag{B.1}$$

along with

$$H(\boldsymbol{\theta},\boldsymbol{\theta}^*) = -\int \log p(\mathbf{s}|\mathbf{y},\boldsymbol{\theta})p(\mathbf{s}|\mathbf{y},\boldsymbol{\theta}^*)d\mathbf{s}. \tag{B.2}$$

Using equation (B.1) and (B.2), we can derive that

$$Z(\boldsymbol{\theta},\boldsymbol{\theta}^*) = l(\mathbf{y}|\boldsymbol{\theta}) - H(\boldsymbol{\theta},\boldsymbol{\theta}^*). \tag{B.3}$$

Since $H(\boldsymbol{\theta},\boldsymbol{\theta}^*)$ is minimized as a function of $\boldsymbol{\theta}$ at $\boldsymbol{\theta}^*$, we have

$$\frac{\partial}{\partial\boldsymbol{\theta}}H(\boldsymbol{\theta},\boldsymbol{\theta}^*)|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = 0. \tag{B.4}$$

Therefore, we have the identity that

$$\frac{\partial}{\partial \boldsymbol{\theta}} l(\mathbf{y}|\boldsymbol{\theta}^*) = \frac{\partial}{\partial \boldsymbol{\theta}} Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*)|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}. \tag{B.5}$$

The intuition behind the Fisher's identity is the fact that $Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ minorizes $l(\mathbf{y}|\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, which makes their derivatives equal provided that both functions are differentiable.

Therefore,

$$\begin{aligned}
Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) &= E_{\boldsymbol{\theta}^*}\left\{\log[u(\mathbf{s}, \mathbf{y}|\boldsymbol{\theta})]|\mathscr{Y}\right\} \\
&= E_{\boldsymbol{\theta}^*}\left\{\log P(\mathbf{s}|\mathscr{Y})\right\} + E_{\boldsymbol{\theta}^*}\left\{\log P(\mathbf{y}|\mathbf{s})|\mathscr{Y}\right\} \\
&= \left[\sum_{m=1}^{M} \gamma^{T_1}(m) \log \pi_m\right] + \sum_{k=1}^{K} \sum_{m=1}^{M} \gamma^{T_k}(m) \log g_m[Y(t); \mathbf{x}_t, \boldsymbol{\theta}_m] \\
&\quad + \sum_{m=1}^{M} \left\{\sum_{\ell \neq m} \left[\log q(m,\ell) E_{\boldsymbol{\theta}^*}(n_{m\ell}|\mathscr{Y})\right] - q(m,m) E_{\boldsymbol{\theta}^*}(\nu_m|\mathscr{Y})\right\},
\end{aligned} \tag{B.6}$$

where $\gamma^t(m)$ is the state probability derived from the forward-backward algorithm; $E_{\boldsymbol{\theta}^*}(n_{m\ell}|\mathscr{Y})$ is the conditional expectation for the number of transitions from state m to state ℓ, and $E_{\boldsymbol{\theta}^*}(\nu_m|\mathscr{Y})$ is the conditional expectation for the total duration in state m, both of which can be computed using the efficient Expm algorithm proposed in [Liu15].

Finally, the gradients of the working parameters can be derived using the chain rule,

$$
\begin{aligned}
\frac{\partial l}{\partial a_m} =& \gamma^{T_1}(m) - \pi_m, \quad m = 1, \dots, M-1, \\
\frac{\partial l}{\partial c_{m,\ell}} =& [1 - q(m,\ell)][E_{\boldsymbol{\theta}^*}(n_{m\ell}|\mathscr{Y}) - E_{\boldsymbol{\theta}^*}(\nu_m|\mathscr{Y})q(m,\ell)], \quad m = 1, \dots, M, \ \ell \neq m, \\
\frac{\partial l}{\partial b_{0,0}} =& -\sum_{t=t_1}^{t_T} \gamma^t(1) \Big\{ \delta[\mathbf{X}(t)] \mathscr{I}[Y(t) > 0] + \\
& \frac{\delta[\mathbf{X}(t)]\{1 - \delta[\mathbf{X}(t)]\}\{e^(-\lambda_1[\mathbf{X}(t)]) - 1\}}{\delta[\mathbf{X}(t)] + \{1 - \delta[\mathbf{X}(t)]\} e^(-\lambda_1[\mathbf{X}(t)])} \mathscr{I}[Y(t) = 0] \Big\}, \\
\frac{\partial l}{\partial b_{1,0}} =& -\sum_{t=t_1}^{t_T} \gamma^t(1) \Big\{ \delta[\mathbf{X}(t)] \mathscr{I}[Y(t) > 0] + \\
& \frac{\delta[\mathbf{X}(t)]\{1 - \delta[\mathbf{X}(t)]\}\{e^(-\lambda_1[\mathbf{X}(t)]) - 1\}}{\delta[\mathbf{X}(t)] + \{1 - \delta[\mathbf{X}(t)]\} e^(-\lambda_1[\mathbf{X}(t)])} \mathscr{I}[Y(t) = 0] \Big\} \mathbf{X}(t), \\
\frac{\partial l}{\partial b_{0,1}} =& \sum_{t=t_1}^{t_T} \gamma^t(1) \Big\{ [Y(t) - \lambda_1(\mathbf{X}(t))] \mathscr{I}[Y(t) > 0] + \\
& \frac{\{\delta[\mathbf{X}(t)] - 1\} \lambda_1[\mathbf{X}(t)] e^(-\lambda_1[\mathbf{X}(t)])}{\delta[\mathbf{X}(t)] + (1 - \delta[\mathbf{X}(t)]) e^(-\lambda_1[\mathbf{X}(t)])} \mathscr{I}[Y(t) = 0] \Big\}, \\
\frac{\partial l}{\partial b_{1,1}} =& \sum_{t=t_1}^{t_T} \gamma^t(1) \Big\{ [Y(t) - \lambda_1(\mathbf{X}(t))] \mathscr{I}[Y(t) > 0] + \\
& \frac{\{\delta[\mathbf{X}(t)] - 1\} \lambda_1[\mathbf{X}(t)] e^(-\lambda_1[\mathbf{X}(t)])}{\delta[\mathbf{X}(t)] + (1 - \delta[\mathbf{X}(t)]) e^(-\lambda_1[\mathbf{X}(t)])} \mathscr{I}[Y(t) = 0] \Big\} \mathbf{X}(t), \\
\frac{\partial l}{\partial b_{0,m}} =& \sum_{t=t_1}^{t_T} \gamma^t(m) \Big\{ Y(t) - \lambda_m(\mathbf{X}(t)) \Big\}, \quad m = 2, \dots, M, \\
\frac{\partial l}{\partial b_{1,m}} =& \sum_{t=t_1}^{t_T} \gamma_t(m) \Big\{ Y(t) - \lambda_m(\mathbf{X}(t)) \Big\} \mathbf{X}(t), \quad m = 2, \dots, M.
\end{aligned}
$$

B.1.3 Proof for Theorem 3.3.1

We first show the almost sure convergence of the constrained maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$ to the true parameter value $\boldsymbol{\theta}^\star$. This follows from the uniform convergence results of the log likelihood [Dou04] for hidden Markov model, along with the feasibility assumption (A1) and identifiability assumption (A5).

Define the conditional log likelihood function for the subject-specific hidden Markov model given initial state S_i^0 as

$$
\begin{aligned}
l_i^T(\theta_i, S_i^0) &\triangleq \log P(Y_i(T_1), \ldots, Y_i(T_K)|S_i(0), \boldsymbol{\theta}_i) \\
&= \sum_{k=1}^{K} \log P(Y_i(T_k)|Y_i(0), \ldots, Y_i(T_{k-1}), S_i(0), \boldsymbol{\theta}_i),
\end{aligned}
$$

for subject $i = 1, \ldots, n$. By assumption (A2), the observed time process and the generative hidden Markov process are independent and do not have common parameters. Under (A0), (A3), (A4), the uniform convergence of the normalized log likelihood is proved in Proposition 2 in [Dou04], i.e.,

$$
\sup_{\boldsymbol{\theta}_i} \sup_{S_i(0)} |K^{-1} l_i^{T_K}(\boldsymbol{\theta}_i, S_i(0))) - l_i(\boldsymbol{\theta}_\mathbf{i})| \to 0 \; a.s. \text{ as } K \to \infty
$$

for $i = 1, \ldots, n$, where $l_i(\boldsymbol{\theta}_i)$ is a deterministic asymptotic function such that

$$
l_i(\boldsymbol{\theta}_i) \triangleq \lim_{K\to\infty} E_{\theta_i^\star}\Big[\log P(Y_i(T_k)|Y_i(-T_K), \ldots, Y_i(T_{k-1}), S_i(-T_K), \boldsymbol{\theta}_i)\Big].
$$

To show that $l_i(\boldsymbol{\theta}_i)$ is maximized at $\boldsymbol{\theta}_i = \boldsymbol{\theta}_i^\star$ for $i = 1, \ldots, n$,

$$
\begin{aligned}
& l_i(\boldsymbol{\theta}_i) - l_i(\boldsymbol{\theta}_i^\star) \\
&= \lim_{K\to\infty} E_{\boldsymbol{\theta}_i^\star}\Big[\log \frac{P(Y_i(T_k)|Y_i(-T_K), \ldots, Y_i(T_{k-1}), S_i(-T_K), \boldsymbol{\theta}_i)}{P(Y_i(T_k)|Y_i(-T_K), \ldots, Y_i(T_{k-1}), S_i(-T_K), \boldsymbol{\theta}_i^\star)}\Big] \\
&= \lim_{k\to\infty} E_{\theta_i^\star}\Big\{E_{\theta_i^\star}\Big[\log \frac{P(Y_i(T_k)|Y_i(-T_K), \ldots, Y_i(T_{k-1}), S_i(-T_K), \boldsymbol{\theta}_i)}{P(Y_i(T_k)|Y_i(-T_K), \ldots, Y_i(T_{k-1}), S_i(-T_K), \boldsymbol{\theta}_i^\star)} \\
& \Big| Y_i(-T_K), \ldots, Y_i(T_{k-1}), S_i(-T_K), \boldsymbol{\theta}_i^\star\Big]\Big\},
\end{aligned}
$$

where the inner conditional expectation is the Kullback-Leibler divergence and is maximized at $\boldsymbol{\theta}_i = \boldsymbol{\theta}_i^\star$. By the identifiability assumption (A5), $\boldsymbol{\theta}_i^\star$ is the unique maximizer of $l_i(\boldsymbol{\theta}_i)$. Therefore, it follows that

$$
\arg\max_{\boldsymbol{\theta}_i} K^{-1} l_i^{T_K}(\boldsymbol{\theta}_i, S_i(0)) \to \arg\max_{\boldsymbol{\theta}_i} l_i(\boldsymbol{\theta}_i),
$$

with probability 1 as $K \to \infty$.

Since $\arg\max_{\boldsymbol{\theta}_i} T^{-1} l_i^T(\boldsymbol{\theta}_i, S_i^0)$ is the MLE $\hat{\boldsymbol{\theta}}_{i,K}$, we have

$$
\hat{\boldsymbol{\theta}}_{i,K} \to \boldsymbol{\theta}_i^\star \text{ with probability 1 for all } i = 1, \ldots, n.
$$

Thus, the unconstrained problem is minimized at $\boldsymbol{\theta}_n = [\hat{\boldsymbol{\theta}}_{1,K_1}, \ldots, \hat{\boldsymbol{\theta}}_{n,K_n}]$ with the minimum value $\hat{f} = \sum_{i=1}^{n} f_i(\hat{\boldsymbol{\theta}}_i)$. By assumption (A1), the constraint set $\mathscr{C} \triangleq \{\boldsymbol{\theta} \in \hat{}\,; \mathbf{D}\boldsymbol{\theta} = \mathbf{0}\}$ is nonempty. And there exists some $r \in \mathbb{R}$ such that the set $\{\boldsymbol{\theta} \in \mathscr{C}; f(\boldsymbol{\theta}) \leq r\}$ is nonempty and compact. Therefore, by Weierstrass theorem, the optimal set for the constrained problem is nonempty, with the optimal function value $f^{\star} \in [\hat{f}, r]$.

Next, we will show the numerical convergence of the ADMM algorithm in the hierarchical hidden Markov model. The numerical convergence properties discussed in a general ADMM problem setting include:

(i) Residual convergence. $\mathbf{A}\tilde{\boldsymbol{\theta}}_n^{(v)} - \mathbf{B}\tilde{\mathbf{z}}_n^{(v)} \rightarrow 0$ as $v \rightarrow \infty$.

(ii) Dual convergence. $\tilde{\xi}_n^{(v)} \rightarrow \hat{\xi}_n$ as $v \rightarrow \infty$.

(iii) Objective convergence. $f(\tilde{\boldsymbol{\theta}}_n^{(v)}) \rightarrow f(\hat{\boldsymbol{\theta}}_n)$ as $v \rightarrow \infty$.

(iv) Primal convergence. $\tilde{\boldsymbol{\theta}}_n^{(v)} \rightarrow \hat{\boldsymbol{\theta}}_n$, $\tilde{\mathbf{z}}^{(v)} \rightarrow \hat{\mathbf{z}}_n$ as $v \rightarrow \infty$.

In order to guarantee the (i) residual convergence, (ii) objective convergence, and (iii) dual convergence as stated above, we need the following two standard assumptions as stated in [Boy11],

(B1) The objective function f is closed, proper, and convex.
(B2) The unaugmented Lagrangian L_0 has a saddle point.

Furthermore, in order to show the (iv) primal convergence, two additional assumptions are proposed in [Shi14],
(C1) $f(\boldsymbol{\theta})$ is strongly convex,
(C2) $\nabla f(\boldsymbol{\theta})$ is Lipschitz continuous.
In the HCTHMM context, we only need (C1) and (C2) to hold in a local neighborhood of the true parameter values. Thus, the proof for the second part of Theorem 1 boils down to showing how assumptions (A0) - (A6) can be used to establish assumptions (B1), (B2), (C1), and (C2).

By assumptions (A0) and (A1), the objective negative log likelihood function $f(.)$ is closed and proper. Together with assumption (A6), we can see that $f(.)$ is a locally convex, closed, proper function. Thus, assumption (B1) holds in the neighborhood of the true parameter value.

To establish assumption (B2), we need to resort to the saddle point theorem (Lagrange, 1797), which indicates that if $\boldsymbol{\theta}$ is a solution to the constrained optimization problem, then there exists $\mathbf{d} \geq 0$ such that $(\boldsymbol{\theta}, \mathbf{d})$ is a saddle point the Lagrangian $L_0(.)$. Since we already showed the feasibility of the constrained maximum likelihood estimation in the first part of the proof under (A0) - (A5), assumption (B2) holds.

As (B1) and (B2) are established above, we can claim the (i) residual convergence, (ii) objective convergence, and (iii) dual convergence in the ADMM updates. To further claim the (iv) primal convergence, we will prove that both (C1) and (C2) hold given assumption (A6). By assumption (A6), for all $\|\boldsymbol{\theta}-\boldsymbol{\theta}^{\star}\|_2 \leq \epsilon$, the eigenvalues of $\nabla^2 f(\theta)$ are between ϱ_- and ϱ_+ where $0 < \varrho_- \leq \varrho_+ < \infty$, such that

$$\|\nabla^2 f(\boldsymbol{\theta})\|_{op} \triangleq \sup_{\|v\|_2=1} \|\nabla^2 f(\boldsymbol{\theta})\mathbf{v}\|_2 \leq \varrho_+.$$

By Taylor series expansion on the gradient of the log negative log likelihood function,

$$\nabla f(\boldsymbol{\theta}) = \nabla f(\boldsymbol{\theta}^{\star}) + \nabla^2 f(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\boldsymbol{\theta}^{\star}),$$

where $\tilde{\boldsymbol{\theta}}$ is between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^{\star}$. Then

$$\begin{aligned}
\|\nabla f(\boldsymbol{\theta})-\nabla f(\boldsymbol{\theta}^{\star})\|_2 &= \|\nabla^2 f(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\boldsymbol{\theta}^{\star})\|_2 \\
&= \|\nabla^2 f(\tilde{\boldsymbol{\theta}})\frac{\boldsymbol{\theta}-\boldsymbol{\theta}^{\star}}{\|\boldsymbol{\theta}-\boldsymbol{\theta}^{\star}\|_2}\|_2 \|\boldsymbol{\theta}-\boldsymbol{\theta}^{\star}\|_2 \\
&\leq \|\nabla^2 f(\tilde{\boldsymbol{\theta}})\|_{op} \|\boldsymbol{\theta}-\boldsymbol{\theta}^{\star}\|_2 \\
&\leq \varrho_+ \|\boldsymbol{\theta}-\boldsymbol{\theta}^{\star}\|_2.
\end{aligned}$$

Therefore, the gradient $\nabla f(\boldsymbol{\theta})$ is Lipschitz continuous in a neighborhood of $\boldsymbol{\theta}^{\star}$ so that (C1) holds.

Also by Taylor expansion on the negative log likelihood function,

$$\begin{aligned}
f(\boldsymbol{\theta}) &= f(\boldsymbol{\theta}^{*}) + \nabla f(\boldsymbol{\theta}^{\star})(\boldsymbol{\theta}-\boldsymbol{\theta}^{\star}) + (\boldsymbol{\theta}-\boldsymbol{\theta}^{\star})'\nabla^2 f(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\boldsymbol{\theta}^{\star}) \\
&\geq f(\boldsymbol{\theta}^{\star}) + \nabla f(\boldsymbol{\theta}^{\star})(\boldsymbol{\theta}-\boldsymbol{\theta}^{\star}) + \varrho_- \|\boldsymbol{\theta}-\boldsymbol{\theta}^{\star}\|_2,
\end{aligned}$$

so that (C2) holds, i.e., f is strongly convex in the neighborhood of $\boldsymbol{\theta}^{\star}$.

B.1.4 Proof for Theorem 3.3.2

The estimated mean probability of state m among subjects in group j is defined as

$$\hat{\phi}_{j,n}(m;\hat{\boldsymbol{\theta}}_n) = \frac{1}{n_j}\sum_{\{i:G_i=j\}} \hat{\eta}_{i,K_i}(m;\boldsymbol{\theta}_i),\ j=1,\ldots,J, \tag{B.7}$$

$$\text{where } \hat{\eta}_{i,K_i}(m;\hat{\boldsymbol{\theta}}_n) = \frac{1}{K_i}\sum_{k=1}^{K_i}\gamma_i^{T_k}(m;\hat{\boldsymbol{\theta}}_n),\ m=1,\ldots,M.$$

By assumptions (A0) - (A5) for a well-defined stationary continuous-time hidden Markov model, the limiting state distribution exists and is determined by the transition kernel as a function of $\boldsymbol{\theta}$. In fact, $\mu_j(m;\hat{\boldsymbol{\theta}}_n) \triangleq E[\hat{\eta}_{i,K_i}(m;\hat{\boldsymbol{\theta}}_n)]$ is the limiting probability of state m for subjects in group j, which is a continuous function of $\hat{\boldsymbol{\theta}}_n$ bounded between 0 and 1. Since $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}^\star$ by Theorem 1, we have $\mu_j(m;\hat{\boldsymbol{\theta}}_n) \xrightarrow{a.s.} \mu_j(m;\boldsymbol{\theta}^\star)$ as $K_i \to \infty$ for $i=1,\ldots,n$, $n_j \to \infty$ for $j=1,\ldots,J$.

Since the subjects are independent with the same transition rate matrix within group, then $\hat{\phi}_{j,n}(m;\hat{\boldsymbol{\theta}}_n) \xrightarrow{a.s.} \mu_j(m;\hat{\boldsymbol{\theta}}_n)$ as $n_j \to \infty$ by Law of Large Numbers. Thus, $\hat{\phi}_{j,n}(m;\hat{\boldsymbol{\theta}}_n) \xrightarrow{a.s.} \mu_j(m;\boldsymbol{\theta}^\star)$ as $T_i \to \infty$ and $n_j \to \infty$.

Furthermore, by the Lindeberg-Lévy Central Limit Theorem,

$$\frac{\hat{\phi}_{j,n}(m;\hat{\boldsymbol{\theta}}_n) - \mu_j(m;\hat{\boldsymbol{\theta}}_n)}{\sqrt{\sigma_j^2(m;\hat{\boldsymbol{\theta}}_n)/n_j}} \xrightarrow{d} \mathcal{N}(0,1),$$

as $K_i \to \infty$ for $i=1,\ldots,n$, $n_j \to \infty$ for $j=1,\ldots,J$.

B.2 Additional information on 03 - 04 NHANES

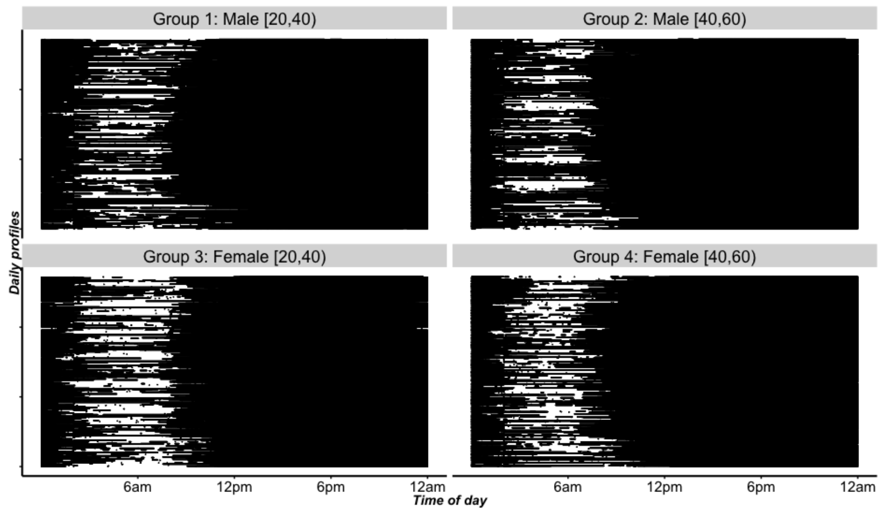

Figure B.1 The missingness (in white) pattern for each daily profile by subgroup.

Figure B.1 shows the missingness pattern in each subgroup, where the majority of the missingness happens during the sleep time (midnight to early morning). There are still sporadic missingness during other periods of time in the day, which may correspond to activities like swimming or bathing. Among those subjects, the average proportion of zeros after accounting for the missingness is around 25%, so that zero-inflation is still an issue to be considered in the modeling.

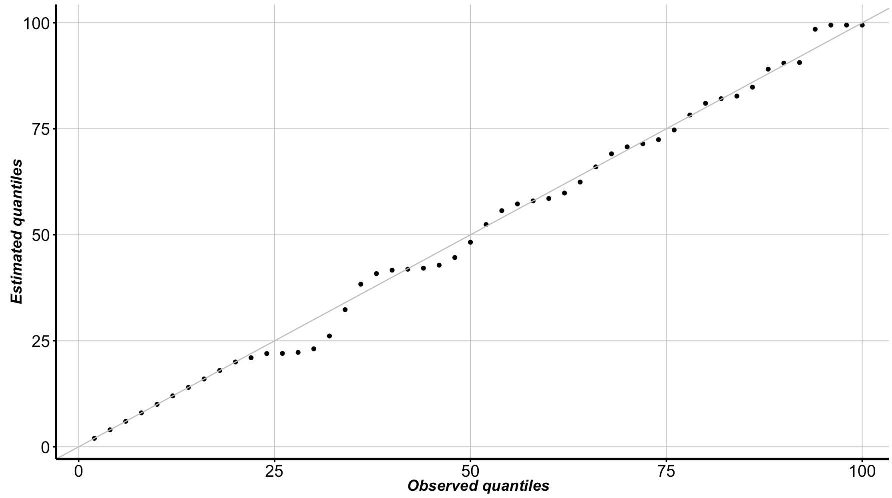

Figure B.2 The estimated quantiles using the fitted model versus the observed quantiles.

Figure B.2 shows a QQ-plot of the estimated quantiles using the fitted hierarchical continuous-time hidden Markov model and the observed quantiles. It can be seen that they follow each other closely and that there are no indications of serious lack-of-fit.

APPENDIX

C

SUPPLEMENTARY MATERIALS FOR CHAPTER 4

C.1 Technical details and proofs

C.2 Proof for Lemma 4.2.1

Since U^{j+1} depends on $\mathbf{H}^{j+1}$ through $(\mathbf{S}^{j+1}, A^{j+1})$, where A^{j+1} depends on $\mathbf{S}^{j+1}$, it is sufficient to show that $\mathbf{S}^{j+1}$ depends on $(\mathbf{H}^j, A^j)$ only through $(\mathbf{S}^j, A^j)$. Since $\mathbf{X}^{j+1}$ depends only on $(\mathbf{B}^{j+1}, \mathbf{X}^j)$ in a Gaussian autoregressive hidden Markov model, then the proof boils down to showing $\mathbf{B}^{j+1}$ depends on $(\mathbf{H}^j, A^j)$ through $(\mathbf{S}^j, A^j)$, where $\mathbf{S}^j \triangleq [\mathbf{B}^j, \mathbf{X}^j]$.

By definition of belief states,

$$\begin{aligned} B_m^{j+1} &\triangleq P\{M(T^{j+1}) = m | A^j, \mathbf{H}^j\} \\ &= \frac{\sum_k P\{M(T^{j+1}) = m, M(T^j) = k | A^j, \mathbf{H}^j\}}{\sum_\ell \sum_k P(M(T^{j+1}) = \ell, M(T^j) = k | A^j, \mathbf{H}^j\}}. \end{aligned}$$

By iterative conditioning, we can show that

$$
\begin{aligned}
&P\{M(T^{j+1})=m, M(T^j)=k|A^j, \mathbf{H}^j\}\\
=&P\{M(T^{j+1})=m|M(T^j)=k, A^j\}P\{M(T^j)=k|A^{j-1}, \mathbf{H}^{j-1}\}\\
=&P\{M(T^{j+1})=m|M(T^j)=k, A^j\}B_k^j,
\end{aligned}
$$

where the distribution of $M(T^j)$ is summarized in $\mathbf{B}^j$. Thus, we've shown $\mathbf{B}^{j+1}$ depends on $(\mathbf{H}^j, A^j)$ through $(\mathbf{B}^j, A^j)$, which means $\mathbf{S}^{j+1}$ depends on $(\mathbf{H}^j, A^j)$ only through $(\mathbf{S}^j, A^j)$.

C.3 Comments on Remark 4.2.1

Remark 4.2.1 states that there exists a generative model such that any deterministic function $\mathbf{R}^j = g(\mathbf{S}^j)$ with $||\text{Cov}(\mathbf{S}^j|\mathbf{R}^j)|| > 0$ w.p.1. implies $\sup_{\pi_g\in\Pi_g} V(\pi_g) < \sup_{\pi\in\Pi} V(\pi)$, where $\Pi_g = \{\pi_g : \text{Supp}\mathbf{R}^j \to \mathscr{A}^j\}$ and $\Pi = \{\pi : \text{Supp}\ \mathbf{S}^j \to \mathscr{A}^j\}$.

To construct such an example, consider a two-state MDP with a binary action and binary reward. Let the state of the MDP be $S^j = X^j \in \{0,1\}$. Suppose that the transition dynamics are $P(S^{j+1} = 1|S^j = s, A^j = a) = 0.5$ for all $(s,a) \in \{0,1\}^2$ and the utility, $U(s,a)$ satisfies: $U(0,0) = 1$, $U(0,1) = 0$, $U(1,0) = 0$, and $U(1,1) = 1$ so that the optimal regime is $\pi^{\text{opt}}(s) \equiv s$. Any deterministic function $g(s)$ such that $\text{Var}\left\{S^j|g(S^j)\right\} > 0$ w.p.1. must be constant and thus any deterministic regime based on R^j has value $V(\pi^{\text{opt}})/2$.

C.4 Proof and comments for Lemma 4.5.1

Under (A1) - (A2), (B1) - (B8), the asymptotic normality of $\widehat{\boldsymbol{\varrho}}_n$ has been shown for general hidden Markov models [Bic98; JP99; LGM00; DM01; Dou04], and in particular for the Gaussian autoregressive hidden Markov model in Theorem 4 in [Dou04]. Since $\mathbf{b}^j(\mathbf{h}^j; \boldsymbol{\varrho})$ is a bounded, continuous function in $\boldsymbol{\varrho}$ for each $\mathbf{h}^j \in \text{dom}\,\mathbf{H}^j$, $j = 1, \ldots, J$, we have point-wise convergence results for $\mathbf{b}^j(\mathbf{h}^j; \widehat{\boldsymbol{\varrho}}_n)$ and $\sqrt{n}\{\mathbf{b}^j(\mathbf{h}^j; \widehat{\boldsymbol{\varrho}}_n) - \mathbf{b}^j(\mathbf{h}^j; \boldsymbol{\varrho}^*)\}$ as $n \to \infty$ for each $\mathbf{h}^j \in \text{dom}\,\mathbf{H}^j$. Along with Assumption (B9),

(B9) For any $\boldsymbol{\varrho}_1, \boldsymbol{\varrho}_2 \in \Theta \times \Gamma$, $||\mathbf{b}^j(\mathbf{h}^j, \boldsymbol{\varrho}_1) - \mathbf{b}^j(\mathbf{h}^j, \boldsymbol{\varrho}_2)|| \le g^j(\mathbf{h}^j)||\boldsymbol{\varrho}_1 - \boldsymbol{\varrho}_2||$, for some integrable function $g : \text{dom}\,\mathbf{H}^j \to \mathbb{R}$, $j \ge 1$;

which is equivalent to a Donsker class condition, so the process convergence results immediately follow. Therefore, the key point here is to show Assumption (B9) holds in the Gaussian autoregressive hidden Markov model. Since $\mathbf{b}^j(\mathbf{h}^j, \boldsymbol{\varrho})$ is differentiable in $\boldsymbol{\varrho}$, it is sufficient to show that each component in the derivative of $\mathbf{b}^j(\mathbf{h}^j, \boldsymbol{\varrho})$ with respect to $\boldsymbol{\varrho}$ is integrable. Then at each fixed $j = 1, \ldots, J$, for any $\boldsymbol{\varrho}_1, \boldsymbol{\varrho}_2 \in \Theta \times \Gamma$,

$$\|\mathbf{b}^j(\mathbf{h}^j, \boldsymbol{\varrho}_1) - \mathbf{b}^j(\mathbf{h}^j, \boldsymbol{\varrho}_2)\| \le \sup_{\tilde{\varrho} \in \Theta \times \Gamma} \left\| \frac{\partial \mathbf{b}^j(\mathbf{h}^j, \tilde{\boldsymbol{\varrho}})}{\partial \boldsymbol{\varrho}} \right\| \left\| \boldsymbol{\varrho}_1 - \boldsymbol{\varrho}_2 \right\|.$$

In a Gaussian autoregressive hidden Markov model with K latent states, we have

$$\boldsymbol{\varrho} = ([q_{k,\ell,a}]_{k,\ell=1,\ldots,K; a \in \mathscr{A}}, [\boldsymbol{\mu}_k]_{k=1}^K, \mathrm{vec}[\Sigma_k]_{k=1}^K, \mathrm{vec}[\Psi_k]_{k=1}^K),$$

which consists of components in the transition kernel for each treatment, along with mean, covariance, and autoregression coefficients in the state-dependent conditional multivariate normal distributions.

We show the proof by induction. Assume some fixed initial state probability vector $[b_1^0, \ldots, b_K^0]$. For convenience, define the operator:

$$f_{\ell,k}(a, \delta, b', \mathbf{x}', \mathbf{x}) \triangleq b' \exp\{\delta Q(a)\}_{\ell,k} |\Sigma_k|^{-1/2} \exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k - \Psi_k \mathbf{x}')^\intercal \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k - \Psi_k \mathbf{x}')\},$$

where $Q(a) \triangleq \{q_{k,\ell,a}\}_{k,\ell=1,\ldots,K}$.

Then at decision point $j = 1$, for each $k = 1, \ldots, K$,

$$b_k^1(\mathbf{h}^1, \boldsymbol{\varrho}) \propto \sum_{\ell=1}^K f_{\ell,k}(A^1, \Delta^1, b_\ell^0, \mathbf{0}, \mathbf{X}^1)$$

To show $\|\frac{\partial \mathbf{b}^1(\mathbf{h}^1, \boldsymbol{\varrho})}{\partial \boldsymbol{\varrho}}\|$ is integrable, it is sufficient to show $\|\frac{\partial f_{\ell,k}(A^1, \Delta^1, b_\ell^0, \mathbf{0}, \mathbf{X}^1)}{\partial q_{k,\ell,a}}\|$, $\|\frac{\partial f_{\ell,k}(A^1, \Delta^1, b_\ell^0, \mathbf{0}, \mathbf{X}^1)}{\partial \boldsymbol{\mu}_k}\|$, $\|\frac{\partial f_{\ell,k}(A^1, \Delta^1, b_\ell^0, \mathbf{0}, \mathbf{X}^1)}{\partial \Sigma_k}\|$, $\|\frac{\partial f_{\ell,k}(A^1, \Delta^1, b_\ell^0, \mathbf{0}, \mathbf{X}^1)}{\partial \Psi_k}\|$ are integrable for each $k, \ell = 1, \ldots, K$, $a \in \mathscr{A}$.

$$\left\| \frac{\partial f_{\ell,k}(A^1, \Delta^1, b_\ell^0, \mathbf{0}, \mathbf{X}^1)}{\partial q_{k,\ell,a}} \right\| = \left\| b_\ell^0 \Delta^1 \exp\{\Delta^1 Q(A^1)\}_{\ell,k} |\Sigma_k|^{-1/2} \exp\{-\frac{1}{2}(\mathbf{X}^1 - \boldsymbol{\mu}_k)^\intercal \Sigma_k^{-1} (\mathbf{X}^1 - \boldsymbol{\mu}_k)\} \right\|,$$

where $b_\ell^0 \Delta^1 \exp\{\Delta^1 Q(A^1)\}_{\ell,k} |\Sigma_k|^{-1/2}$ is bounded because $\Delta^1 \in (0, 1]$, Σ_k is a positive definite covariance matrix, and $\boldsymbol{\varrho}$ has a compact support by Assumption (B2); also, $\exp\{-\frac{1}{2}(\mathbf{X}^1 - \boldsymbol{\mu}_k)^\intercal \Sigma_k^{-1} (\mathbf{X}^1 - \boldsymbol{\mu}_k)\}$ is bounded between $[0, 1]$. Thus, $\sup_{\varrho \in \Theta \times \Gamma} \|\frac{\partial f_{\ell,k}(A^1, \Delta^1, b_\ell^0, \mathbf{0}, \mathbf{X}^1)}{\partial q_{k,\ell,a}}\|$ is integrable.

Next,

$$\left\|\frac{\partial f_{\ell,k}(A^1,\Delta^1,b_\ell^0,\mathbf{0},\mathbf{X}^1)}{\partial \boldsymbol{\mu}_k}\right\| = \Big\| b_\ell^0 \exp\{\Delta^1 Q(A^1)\}_{\ell,k}|\Sigma_k|^{-1/2}\exp\{-\frac{1}{2}(\mathbf{X}^1-\boldsymbol{\mu}_k)^\intercal\Sigma_k^{-1}(\mathbf{X}^1-\boldsymbol{\mu}_k)\}$$
$$\Sigma_k^{-1}(\mathbf{X}^1-\boldsymbol{\mu}_k)\Big\|$$
$$\leq C_1\left\|\Sigma_k^{-1}(\mathbf{X}^1-\boldsymbol{\mu}_k)\right\|,$$

for some positive constant C_1. Since $\mathbf{X}^1$ follows a multivariate normal distribution, $\|\Sigma_k^{-1}(\mathbf{X}^1-\boldsymbol{\mu}_k)\|$ is integrable. Thus, $\sup_{\varrho\in\Theta\times\Gamma}\|\frac{\partial f_{\ell,k}(A^1,\Delta^1,b_\ell^0,\mathbf{0},\mathbf{X}^1)}{\partial \boldsymbol{\mu}_k}\|$ is integrable. Next,

$$\left\|\frac{\partial f_{\ell,k}(A^1,\Delta^1,b_\ell^0,\mathbf{0},\mathbf{X}^1)}{\partial \Psi_k}\right\| = \mathbf{0},$$
$$\left\|\frac{\partial f_{\ell,k}(A^1,\Delta^1,b_\ell^0,\mathbf{0},\mathbf{X}^1)}{\partial \Sigma_k}\right\| = \Big\|\frac{1}{2} b_\ell^0 \exp\{\Delta^1 Q(A^1)\}_{\ell,k}|\Sigma_k|^{-1/2}\exp\{-\frac{1}{2}(\mathbf{X}^1-\boldsymbol{\mu}_k)^\intercal\Sigma_k^{-1}(\mathbf{X}^1-\boldsymbol{\mu}_k)\}$$
$$\{\Sigma_k^{-1}-\Sigma_k^{-1}(\mathbf{X}^1-\boldsymbol{\mu}_k)(\mathbf{X}^1-\boldsymbol{\mu}_k)^\intercal\Sigma_k^{-1}\}\Big\|$$
$$\leq C_2\left\|\Sigma_k^{-1}-\Sigma_k^{-1}(\mathbf{X}^1-\boldsymbol{\mu}_k)(\mathbf{X}^1-\boldsymbol{\mu}_k)^\intercal\Sigma_k^{-1}\right\|$$

where $\|\Sigma_k^{-1}-\Sigma_k^{-1}(\mathbf{X}^1-\boldsymbol{\mu}_k)(\mathbf{X}^1-\boldsymbol{\mu}_k)^\intercal\Sigma_k^{-1}\|$ is integrable since $\mathbf{X}^1$ follows a multivariate normal distribution. Putting things together, we can define $g^1(\mathbf{h}^1)=\sup_{\varrho\in\Theta\times\Gamma}\|\frac{\partial \mathbf{b}^1(\mathbf{h}^1,\boldsymbol{\varrho})}{\partial \boldsymbol{\varrho}}\|$, such that for any $\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2\in\Theta\times\Gamma$, $\|\mathbf{b}^1(\mathbf{h}^1,\boldsymbol{\varrho}_1)-\mathbf{b}^1(\mathbf{h}^1,\boldsymbol{\varrho}_2)\|\leq g^1(\mathbf{h}^1)\|\boldsymbol{\varrho}_1-\boldsymbol{\varrho}_2\|$.

At decision point $j=2$, the component-wise derivatives are

$$\left\|\frac{\partial f_{\ell,k}(A^2,\Delta^2,b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho}),\mathbf{X}^1,\mathbf{X}^2)}{\partial q_{k,\ell,a}}\right\| = \Big\|\frac{\partial b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})}{\partial q_{k,\ell,a}}\exp\{\Delta^2 Q(A^2)\}_{\ell,k}|\Sigma_k|^{-1/2}$$
$$\exp\{-\frac{1}{2}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)^\intercal\Sigma_k^{-1}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)\}+$$
$$b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})\Delta^2\exp\{\Delta^2 Q(A^2)\}_{\ell,k}|\Sigma_k|^{-1/2}$$
$$\exp\{-\frac{1}{2}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)^\intercal\Sigma_k^{-1}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)\}\Big\|$$
$$\leq C_3\left\|\frac{\partial b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})}{\partial q_{k,\ell,a}}\right\|+C_4,$$

for some positive constants C_3, C_4, which is integrable since $\left\|\frac{\partial b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})}{\partial q_{k,\ell,a}}\right\|$ is integrable. By the same arguments,

$$\begin{aligned}\left\|\frac{\partial f_{\ell,k}(A^2,\Delta^2,b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho}),\mathbf{X}^1,\mathbf{X}^2)}{\partial\boldsymbol{\mu}_k}\right\| =& \Bigg\|\frac{\partial b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})}{\partial\boldsymbol{\mu}_k}\exp\{\Delta^2 Q(A^2)\}_{\ell,k}|\Sigma_k|^{-1/2}\\ &\exp\{-\frac{1}{2}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)^\intercal\Sigma_k^{-1}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)\}+\\ &b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})\exp\{\Delta^2 Q(A^2)\}_{\ell,k}|\Sigma_k|^{-1/2}\\ &\exp\{-\frac{1}{2}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)^\intercal\Sigma_k^{-1}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)\}\\ &\Sigma_k^{-1}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)\Bigg\|\\ \le& C_5\left\|\frac{\partial b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})}{\partial\boldsymbol{\mu}_k}\right\|+C_6\left\|\Sigma_k^{-1}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)\right\|,\end{aligned}$$

for some positive constants C_5, C_6, where $\left\|\frac{\partial b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})}{\partial\boldsymbol{\mu}_k}\right\|$ is integrable and $\left\|\Sigma_k^{-1}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)\right\|$ is integrable because $(\mathbf{X}^1,\mathbf{X}^2)$ follows a multivariate normal distribution. Next, we have similar integrability results for

$$\begin{aligned}\left\|\frac{\partial f_{\ell,k}(A^2,\Delta^2,b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho}),\mathbf{X}^1,\mathbf{X}^2)}{\partial\Sigma_k}\right\| =& \Bigg\|\frac{\partial b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})}{\partial\Sigma_k}\exp\{\Delta^2 Q(A^2)\}_{\ell,k}|\Sigma_k|^{-1/2}\\ &\exp\{-\frac{1}{2}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)^\intercal\Sigma_k^{-1}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)\}+\\ &b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})\exp\{\Delta^2 Q(A^2)\}_{\ell,k}|\Sigma_k|^{-1/2}\\ &\exp\{-\frac{1}{2}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)^\intercal\Sigma_k^{-1}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)\}\\ &\{\Sigma_k^{-1}-\Sigma_k^{-1}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)^\intercal\Sigma_k^{-1}\}\Bigg\|\\ \le& C_7\left\|\frac{\partial b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})}{\partial\Sigma_k}\right\|+\\ &C_8\left\|\Sigma_k^{-1}-\Sigma_k^{-1}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)^\intercal\Sigma_k^{-1}\right\|,\end{aligned}$$

and

$$
\begin{aligned}
\left\|\frac{\partial f_{\ell,k}(A^2,\Delta^2,b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho}),\mathbf{X}^1,\mathbf{X}^2)}{\partial\Psi_k}\right\| =& \Bigg\|\frac{\partial b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})}{\partial\Psi_k}\exp\{\Delta^2 Q(A^2)\}_{\ell,k}|\Sigma_k|^{-1/2}\\
&\exp\{-\frac{1}{2}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)^\intercal\Sigma_k^{-1}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)\}+\\
&b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})\exp\{\Delta^2 Q(A^2)\}_{\ell,k}|\Sigma_k|^{-1/2}\\
&\exp\{-\frac{1}{2}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)^\intercal\Sigma_k^{-1}(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)\}\\
&\Sigma_k^{-1}\mathbf{X}^1(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)\Bigg\|\\
\le& C_9\left\|\frac{\partial b_\ell^1(\mathbf{h}^1,\boldsymbol{\varrho})}{\partial\Psi_k}\right\|+C_{10}\left\|\Sigma_k^{-1}\mathbf{X}^1(\mathbf{X}^2-\boldsymbol{\mu}_k-\Psi_k\mathbf{X}^1)\right\|
\end{aligned}
$$

for some positive constants C_7, C_8, C_9, C_{10}. So, we can define $g^2(\mathbf{h}^2)=\sup_{\varrho\in\Theta\times\Gamma}\|\frac{\partial\mathbf{b}^2(\mathbf{h}^2,\boldsymbol{\varrho})}{\partial\boldsymbol{\varrho}}\|$, such that for any $\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2\in\Theta\times\Gamma$, $\|\mathbf{b}^2(\mathbf{h}^2,\boldsymbol{\varrho}_1)-\mathbf{b}^2(\mathbf{h}^2,\boldsymbol{\varrho}_2)\|\le g^2(\mathbf{h}^2)\|\boldsymbol{\varrho}_1-\boldsymbol{\varrho}_2\|$.

Suppose Assumption (B9) holds for decision point j, then it will also hold for decision point $j+1$ using the same methodology shown for decision point $j=2$ since $\Delta^{j+1}\in(0,1]$, and $(\mathbf{X}^j,\mathbf{X}^{j+1})$ follows a multivariate normal distribution. Therefore, we can define $g^j(\mathbf{h}^j)=\sup_{\varrho\in\Theta\times\Gamma}\|\frac{\partial\mathbf{b}^j(\mathbf{h}^j,\boldsymbol{\varrho})}{\partial\boldsymbol{\varrho}}\|$, such that for any $\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2\in\Theta\times\Gamma$, $\|\mathbf{b}^j(\mathbf{h}^j,\boldsymbol{\varrho}_1)-\mathbf{b}^j(\mathbf{h}^j,\boldsymbol{\varrho}_2)\|\le g^j(\mathbf{h}^j)\|\boldsymbol{\varrho}_1-\boldsymbol{\varrho}_2\|$.

C.5 Proof of Theorems 4.5.1 and 4.5.2

To show $\widehat{\boldsymbol{\alpha}}_n(\pi)\xrightarrow{p}\boldsymbol{\alpha}^*(\pi)$ (pointwise) for each $\pi\in\Pi$ as $n\to\infty$, we first show $\{\Lambda_{\text{dis}}(\pi,\boldsymbol{\alpha}):\boldsymbol{\alpha}\in\Omega\}$ is a Glivenko-Cantelli class. From Assumptions (C6), (C7), and the Cauchy-Schwarz inequality, it follows that

$$
\begin{aligned}
\left\|\Lambda_{\text{dis}}(\pi,\boldsymbol{\alpha}_2)-\Lambda_{\text{dis}}(\pi,\boldsymbol{\alpha}_1)\right\| =& \left\|\mathbb{E}\Big[\sum_{j=1}^{J-1}\frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j|\mathbf{S}^j)}\Big\{\gamma\phi(\mathbf{S}^{j+1})-\phi(\mathbf{S}^j)\Big\}^\intercal(\boldsymbol{\alpha}_2-\boldsymbol{\alpha}_1)\phi(\mathbf{S}^j)\Big]\right\|\\
\le& \mathbb{E}\Big[\left\|\sum_{j=1}^{J-1}\frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j|\mathbf{S}^j)}\Big\{\gamma\phi(\mathbf{S}^{j+1})-\phi(\mathbf{S}^j)\Big\}\right\|\left\|\phi(\mathbf{S}^j)\right\|\left\|\boldsymbol{\alpha}_2-\boldsymbol{\alpha}_1\right\|\Big]\\
\le& c_0\|\boldsymbol{\alpha}_2-\boldsymbol{\alpha}_1\|,
\end{aligned}
$$

for some constant $c_0>0$. Thus, $\{\Lambda_{\text{dis}}(\pi,\boldsymbol{\alpha}):\boldsymbol{\alpha}\in\Omega\}$ is a Lipschitz continuous class on compact support and thus Glivenko-Cantelli. Using assumptions (C1), (C2), and Theorem 2.10 of [Kos08], it follows that $\widehat{\boldsymbol{\alpha}}_n(\pi)\xrightarrow{p}\boldsymbol{\alpha}^*(\pi)$ for each $\pi\in\Pi$ as $n\to\infty$.

Using the continuous mapping theorem, $\widehat{\boldsymbol{\alpha}}_n(\pi) \xrightarrow{p} \boldsymbol{\alpha}^*(\pi)$ implies $\int \phi(s)^\intercal \widehat{\boldsymbol{\alpha}}_n(\pi) dR(\mathbf{s}) \xrightarrow{p} \int \phi(s)^\intercal \boldsymbol{\alpha}^*(\pi) dR(\mathbf{s})$, or equivalently, $\widehat{V}_{\text{dis},n}(\pi) \xrightarrow{p} V_{\text{dis}}(\pi)$ for each fixed $\pi \in \Pi$. In order to show $\widehat{V}_{\text{dis},n}(\widehat{\pi}_{\text{dis},n}) \xrightarrow{p} V_{\text{dis}}(\pi^*_{\text{dis}})$, it is sufficient to show $\sup_{\pi\in\boldsymbol{\Pi}} \widehat{V}_{\text{dis},n}(\pi) \xrightarrow{p} \sup_{\pi\in\boldsymbol{\Pi}} V_{\text{dis}}(\pi)$ which follows immediately as $\boldsymbol{\Pi}$ is finite.

We want to derive the limiting distribution of $\sqrt{n}\{\widehat{V}^R_{\text{dis},n}(\pi) - V^R_{\text{dis}}(\pi)\}$ jointly over $\pi \in \Pi$. First, define the auxiliary quantity

$$\widetilde{\Lambda}_{\text{dis},n}(\pi, \boldsymbol{\alpha}) \triangleq \mathbb{P}_n \sum_{j=1}^{J-1} \left[\frac{\pi_{A^j}\left(\mathbf{S}^j\right)}{P\left(A^j \middle| \mathbf{S}^j\right)} \left\{ U^j + \gamma \phi\left(\mathbf{S}^{j+1}\right)^\intercal \boldsymbol{\alpha} - \phi\left(\mathbf{S}^j\right)^\intercal \boldsymbol{\alpha} \right\} \phi\left(\mathbf{S}^j\right) \right];$$

Note that in $\widetilde{\Lambda}_{\text{dis},n}(\pi, \boldsymbol{\alpha})$ differs from $\widehat{\Lambda}_{\text{dis},n}(\pi, \boldsymbol{\alpha})$ in that the true summaries $\mathbf{S}^j$ are used instead of the plug-in estimators $\widehat{\mathbf{S}}^j_n$; it also differs from $\Lambda_{\text{dis}}(\pi, \boldsymbol{\alpha})$ in that empirical measure $\mathbb{P}_n$ is used instead of expectation $\mathbb{E}$. Subsequently, let $\widetilde{\boldsymbol{\alpha}}_n(\pi)$ be a solution to $\widetilde{\Lambda}_{\text{dis},n}(\pi, \boldsymbol{\alpha}) = 0$, and let $\widetilde{V}^R_{\text{dis},n}(\pi) \triangleq [\int \phi(\mathbf{s}) dR(\mathbf{s})]^\intercal \widetilde{\boldsymbol{\alpha}}_n(\pi)$. Let $\Pi = \{\pi_1, \ldots, \pi_L\}$ and define:

$$\begin{aligned}
\widetilde{\mathbf{V}}^R_{\text{dis},n} &\triangleq \{\widetilde{V}^R_{\text{dis},n}(\pi_1), \ldots, \widetilde{V}^R_{\text{dis},n}(\pi_L)\}^\intercal, \\
\mathbf{V}^R_{\text{dis}} &\triangleq \{V^R_{\text{dis}}(\pi_1), \ldots, V^R_{\text{dis}}(\pi_L)\}^\intercal, \\
\widehat{\mathbf{V}}^R_{\text{dis},n} &\triangleq \{\widehat{V}^R_{\text{dis},n}(\pi_1), \ldots, \widehat{V}^R_{\text{dis},n}(\pi_L)\}^\intercal, \\
\widetilde{\boldsymbol{\alpha}}_n &\triangleq \begin{bmatrix} \widetilde{\boldsymbol{\alpha}}_n(\pi_1)^\intercal \\ \vdots \\ \widetilde{\boldsymbol{\alpha}}_n(\pi_L)^\intercal \end{bmatrix} \in \mathbb{R}^{L\times p}, \ \boldsymbol{\alpha}^* \triangleq \begin{bmatrix} \boldsymbol{\alpha}^*(\pi_1)^\intercal \\ \vdots \\ \boldsymbol{\alpha}^*(\pi_L)^\intercal \end{bmatrix} \in \mathbb{R}^{L\times p}, \ \widehat{\boldsymbol{\alpha}}_n \triangleq \begin{bmatrix} \widehat{\boldsymbol{\alpha}}_n(\pi_1)^\intercal \\ \vdots \\ \widehat{\boldsymbol{\alpha}}_n(\pi_L)^\intercal \end{bmatrix} \in \mathbb{R}^{L\times p}.
\end{aligned}$$

Since we can decompose $\sqrt{n}(\widehat{\mathbf{V}}^R_{\text{dis},n} - \mathbf{V}^R_{\text{dis}}) = \sqrt{n}(\widehat{\mathbf{V}}^R_{\text{dis},n} - \widetilde{\mathbf{V}}^R_{\text{dis},n}) + \sqrt{n}(\widetilde{\mathbf{V}}^R_{\text{dis},n} - \mathbf{V}^R_{\text{dis}})$, we will show $\sqrt{n}(\widehat{\mathbf{V}}^R_{\text{dis},n} - \widetilde{\mathbf{V}}^R_{\text{dis},n}) \rightsquigarrow \mathbb{Z}_1$, which is the variability induced by estimating $\widehat{\boldsymbol{\varrho}}_n$, and $\sqrt{n}(\widetilde{\mathbf{V}}^R_{\text{dis},n} - \mathbf{V}^R_{\text{dis}}) \rightsquigarrow \mathbb{Z}_2$, which accounts for the remaining variability independent of estimating $\widehat{\boldsymbol{\varrho}}_n$, where $\mathbb{Z}_1$ and $\mathbb{Z}_2$ are mean zero normal random variables with respective covariances. Then, the finite-dimensional joint distribution of $[\sqrt{n}(\widehat{\mathbf{V}}^R_{\text{dis},n} - \widetilde{\mathbf{V}}^R_{\text{dis},n}), \sqrt{n}(\widetilde{\mathbf{V}}^R_{\text{dis},n} - \mathbf{V}^R_{\text{dis}})]$ is established by Cramer-Wold device, and the final result follows from the continuous mapping theorem and Slutsky's theorem.

First, we study the term $\sqrt{n}(\widehat{\mathbf{V}}^R_{\text{dis},n} - \widetilde{\mathbf{V}}^R_{\text{dis},n}) = \sqrt{n}(\widehat{\boldsymbol{\alpha}}_n - \widetilde{\boldsymbol{\alpha}}_n)^\intercal \int \phi(\mathbf{s}) dR(\mathbf{s})$. Since

$$\begin{aligned}
\widetilde{\boldsymbol{\alpha}}_n(\pi) &= \left(\mathbb{P}_n \left[\sum_{j=1}^{J-1} \frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j|\mathbf{S}^j)} \phi(\mathbf{S}^j) \left\{ \phi(\mathbf{S}^j) - \gamma \phi(\mathbf{S}^{j+1}) \right\}^\intercal \right] \right)^{-1} \mathbb{P}_n \left[\sum_{j=1}^{J-1} \mathbf{G}^j_{\text{dis}}(\pi, \mathbf{H}^j, A^j; \boldsymbol{\varrho}^*) \right], \\
\widehat{\boldsymbol{\alpha}}_n(\pi) &= \left(\mathbb{P}_n \left[\sum_{j=1}^{J-1} \frac{\pi_{A^j}(\widehat{\mathbf{S}}^j_n)}{P(A^j|\widehat{\mathbf{S}}^j_n)} \phi(\widehat{\mathbf{S}}^j_n) \left\{ \phi(\widehat{\mathbf{S}}^j_n) - \gamma \phi(\widehat{\mathbf{S}}^{j+1}_n) \right\}^\intercal \right] \right)^{-1} \mathbb{P}_n \left[\sum_{j=1}^{J-1} \mathbf{G}^j_{\text{dis}}(\pi, \mathbf{H}^j, A^j; \widehat{\boldsymbol{\varrho}}_n) \right].
\end{aligned}$$

Then by Lemma 4.5.1, Assumption (C8), weak law of large numbers and dominated convergence theorem, for each $\pi \in \Pi$,

$$\sqrt{n}\{\widehat{\boldsymbol{\alpha}}_n(\pi) - \widetilde{\boldsymbol{\alpha}}_n(\pi)\} = C_1^{-1}(\pi)C_3(\pi)\sqrt{n}(\widehat{\boldsymbol{\varrho}}_n - \boldsymbol{\varrho}^*) + o_p(1),$$

which implies

$$\sqrt{n}(\widehat{\boldsymbol{\alpha}}_n - \widetilde{\boldsymbol{\alpha}})^\intercal \int \phi(\mathbf{s})dR(\mathbf{s}) = \sqrt{n}\begin{bmatrix} (\widehat{\boldsymbol{\varrho}}_n - \boldsymbol{\varrho}^*)^\intercal C_3(\pi_1)^\intercal C_1(\pi_1)^{-\intercal} \int \phi(\mathbf{s})dR(\mathbf{s}) \\ \vdots \\ (\widehat{\boldsymbol{\varrho}}_n - \boldsymbol{\varrho}^*)^\intercal C_3(\pi_L)^\intercal C_1(\pi_L)^{-\intercal} \int \phi(\mathbf{s})dR(\mathbf{s}) \end{bmatrix} + o_p(1).$$

By the multivariate central limit theorem, $\sqrt{n}(\widehat{\mathbf{V}}^R_{\text{dis},n} - \widetilde{\mathbf{V}}^R_{\text{dis},n}) \rightsquigarrow \mathbb{Z}_1$, where $\mathbb{Z}_1$ is a mean zero Gaussian random variable with covariance Σ_1 whose the $(i,j)^{th}$ element is

$$\left[\int \phi(\mathbf{s})dR(\mathbf{s})\right]^\intercal C_1^{-1}(\pi_i)C_3(\pi_i)\mathbf{I}^{-1}(\boldsymbol{\varrho}^*)C_3(\pi_j)^\intercal C_1^{-\intercal}(\pi_j)\left[\int \phi(\mathbf{s})dR(\mathbf{s})\right].$$

Next, we study the quantity $\sqrt{n}\{\widetilde{V}^R_{\text{dis},n}(\pi) - V^R_{\text{dis}}(\pi)\} = \sqrt{n}(\widetilde{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha})^\intercal \int \phi(\mathbf{s})dR(\mathbf{s})$, where

$$\sqrt{n}\{\widetilde{\boldsymbol{\alpha}}_n(\pi) - \boldsymbol{\alpha}(\pi)\} = C_1^{-1}(\pi)\sqrt{n}(\mathbb{P}_n - P)\left[\sum_{j=1}^{J-1}\mathbf{G}^j_{\text{dis}}(\pi, \mathbf{H}^j, A^j; \boldsymbol{\varrho}^*)\right] + o_p(1),$$

which implies,

$$\sqrt{n}(\widetilde{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha})^\intercal \int \phi(\mathbf{s})dR(\mathbf{s}) = \sqrt{n}(\mathbb{P}_n - P)\begin{bmatrix} \sum_{j=1}^J \mathbf{G}^j_{\text{dis}}(\pi, \mathbf{H}^j, A^j; \boldsymbol{\varrho}^*)^\intercal C_1(\pi_1)^{-\intercal} \int \phi(\mathbf{s})dR(\mathbf{s}) \\ \vdots \\ \sum_{j=1}^J \mathbf{G}^j_{\text{dis}}(\pi, \mathbf{H}^j, A^j; \boldsymbol{\varrho}^*)^\intercal C_1(\pi_L)^{-\intercal} \int \phi(\mathbf{s})dR(\mathbf{s}) \end{bmatrix} + o_p(1).$$

By the multivariate central limit theorem, $\sqrt{n}(\widetilde{\mathbf{V}}^R_{\text{dis},n} - \mathbf{V}^R_{\text{dis}}) \rightsquigarrow \mathbb{Z}_2$, where $\mathbb{Z}_2$ is a mean zero Gaussian random variable with covariance Σ_2 whose the $(i,j)^{th}$ element is

$$\left[\int \phi(\mathbf{s})dR(\mathbf{s})\right]^\intercal C_1^{-1}(\pi_i)C_2(\pi_i, \pi_j)C_1^{-\intercal}(\pi_j)\left[\int \phi(\mathbf{s})dR(\mathbf{s})\right].$$

By the Cramer-Wold device, we can show the finite-dimensional joint convergence: $[\sqrt{n}(\widehat{\mathbf{V}}^R_{\text{dis},n} - \widetilde{\mathbf{V}}^R_{\text{dis},n}), \sqrt{n}(\widetilde{\mathbf{V}}^R_{\text{dis},n} - \mathbf{V}^R_{\text{dis}})] \rightsquigarrow [\mathbb{Z}_1, \mathbb{Z}_2]$. The rest of the theorem follows from the continuous mapping theorem and Slutsky's theorem.

C.5.1 Additional causal assumptions in case study

We review them here only briefly [see Tsi19, for a full development]. Define $\overline{\mathbf{a}}^j \triangleq (a^1, \ldots, a^j) \in \overline{\mathscr{A}}^j$, where $\overline{\mathscr{A}}^j$ is the set of all possible treatment trajectories in the first j visits. Define the set of potential outcomes

$$\mathbf{W}^* \triangleq \left\{ \mathbf{S}^1, \mathbf{S}^{*2}(a^1), \mathbf{S}^{*3}(\overline{\mathbf{a}}^2), \ldots, \mathbf{S}^{*J}(\overline{\mathbf{a}}^{J-1}) \text{ for all } \overline{\mathbf{a}}^{J-1} \in \overline{\mathscr{A}}^{J-1} \right\},$$

where $\mathbf{S}^{*j}(\overline{\mathbf{a}}^{j-1})$ is the potential summary that would have been observed if the patient had followed treatment sequence $\overline{\mathbf{a}}^{j-1}$. By the same token, the potential utility under treatment sequence $\overline{\mathbf{a}}^j$ is defined by $U^{*j}(\overline{\mathbf{a}}^j)$. Accordingly, the potential utility under a regime π is $U^{*j}(\pi) = \sum_{\overline{\mathbf{a}}^j} U^{*j}(\overline{\mathbf{a}}^j) \prod_{v=1}^{j} I[\xi_\pi^v\{\mathbf{S}^{*v}(\overline{\mathbf{a}}^{v-1})\} = a^v]$, where $\{\xi_\pi^j\}_{j\geq 1}$ is a sequence of independent random variables defined on $\mathscr{S}$ such that $\xi_\pi^j(\mathbf{s}^j) \in \mathscr{A}$ and $P\{\xi_\pi^j(\mathbf{s}^j) = a\} \equiv \pi_a(\mathbf{s}^j)$.

Formally, the three causal assumptions are: (Consistency) $\mathbf{S}^j(\overline{\mathbf{A}}^{j-1}) = \mathbf{S}^{*j}(\overline{\mathbf{A}}^{j-1})$ for $j \geq 2$; (Positivity) $P\{A^j = a | \mathbf{S}^j = \mathbf{s}\} > 0$ for all $a \in \mathscr{A}$ and $\mathbf{s} \in \mathscr{S}$; (Ignorability) $\mathbf{W}^* \perp A^j | \mathbf{S}^j$ for $j \geq 2$. Under these assumptions, the state-value function in the discounted utility setting can be written as:

$$\begin{aligned}
v_{dis}(\pi, \mathbf{s}^j) &= \lim_{J\to\infty} \mathbb{E}\left\{ \sum_{v=0}^{K} \gamma^v U^{*(j+v)}(\pi) \middle| \mathbf{S}^j = \mathbf{s}^j \right\} \\
&= \lim_{J\to\infty} \sum_{v=0}^{J} \mathbb{E}\left\{ \gamma^v U^{*(j+v)}(\pi) \middle| \mathbf{S}^j = \mathbf{s}^j \right\} \\
&= \lim_{J\to\infty} \sum_{v=0}^{J} \gamma^v \Bigg\{ \sum_{a^j,\ldots,a^{j+v}} U^{*(j+v)}(\overline{a}^{j-1}, a^j, \ldots, a^{j+v}) \\
&\quad \prod_{w=j}^{j+v} \mathbb{E}\Big(I[\xi_\pi^w\{\mathbf{S}^{*w}(\overline{a}^{w-1})\} = a^w] | \mathbf{S}^j = \mathbf{s}^j \Big) \Bigg\} \\
&= \lim_{J\to\infty} \sum_{v=0}^{J} \gamma^v \sum_{a^j,\ldots,a^{j+v}} U^{*(j+v)}(\overline{a}^{j-1}, a^j, \ldots, a^{j+v}) \\
&\quad \prod_{w=j}^{j+v} \left[\pi_{a^w}\{\mathbf{S}^{*w}(\overline{a}^{w-1})\} \frac{P\{a^w | \mathbf{S}^{*w}(\overline{a}^{w-1})\}}{P\{a^w | \mathbf{S}^{*w}(\overline{a}^{w-1})\}} \right] \\
&= \lim_{J\to\infty} \sum_{v=0}^{J} \gamma^v \mathbb{E}\left[U^{j+v} \prod_{w=j}^{j+v} \left\{ \frac{\pi_{A^w}(\mathbf{S}^w)}{P(A^w | \mathbf{S}^w)} \right\} \middle| \mathbf{S}^j = \mathbf{s}^j \right],
\end{aligned}$$

where the last equality follows from the consistency assumption and the ignorability assumption. Therefore, it follows that

$$\nu_{dis}(\pi, \mathbf{s}^j) = \mathbb{E}\left[\frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j|\mathbf{S}^j)}\left\{U^j + \gamma\,\nu_{dis,\gamma}(\pi, \mathbf{S}^{j+1})\right\}\Big|\mathbf{S}^j = \mathbf{s}^j\right].$$

This indicates that for any function $\phi : \mathscr{S} \to \mathbb{R}^d$, we have

$$0 = \mathbb{E}\left[\frac{\pi_{A^j}(\mathbf{S}^j)}{P(A^j\big|\mathbf{S}^j)}\left\{U^j + \gamma\,\nu_{\text{dis}}(\pi, \mathbf{S}^{j+1}) - \nu_{\text{dis}}(\pi, \mathbf{S}^j)\right\}\phi(\mathbf{S}^j)\right],$$

which is the same estimating equation as equation (1) in Section 3 of the paper. Similarly, we can also show that the estimating equation for the average utility setting is still equation (3) in Section 3 of the paper. Thus, the rest of the theoretical properties of the proposed estimator carries through.

C.6 Tables for medications in STEP-BD

Table C.1 List of common antidepressants in STEP-BD. The dose is divided into 3 levels: low, medium, and high.

Medication name	Low dose (mg)	Medium dose (mg)	High dose (mg)
Deseryl	< 200	$200-400$	> 400
Serzone	< 200	$200-400$	> 400
Citalopram	< 20	$20-40$	> 40
Escitalopram Oxalate	< 10	$10-20$	> 20
Prozac	< 20	$20-40$	> 40
Fluvoxamine	< 100	$100-200$	> 200
Paroxetine	< 20	$20-40$	> 40
Zoloft	< 50	$50-100$	> 100
Venlafaxine	< 75	$75-150$	> 150
Bupropion	< 150	$150-300$	> 300

Table C.2 List of common mood stabilizers in STEP-BD. The dose is divided into 3 levels: low, medium, and high.

Medication name	Low dose (mg)	Medium dose (mg)	High dose (mg)
Tegretol	< 400	$400 - 800$	> 800
Valproate	< 1000	$1000 - 2000$	> 2000
Olanzapine	< 10	$10 - 20$	> 20
Quetiapine	< 400	$400 - 800$	> 800
Clozapine	< 200	$200 - 400$	> 400
Lithium	< 900	$900 - 1800$	> 1800
Risperdal	< 2	$2 - 4$	> 4
Geodon	< 80	$80 - 160$	> 160
Abilify	< 15	$15 - 30$	> 30
Lamictal	< 100	$100 - 200$	> 200

Estimated optimal treatment regime for discounted utility

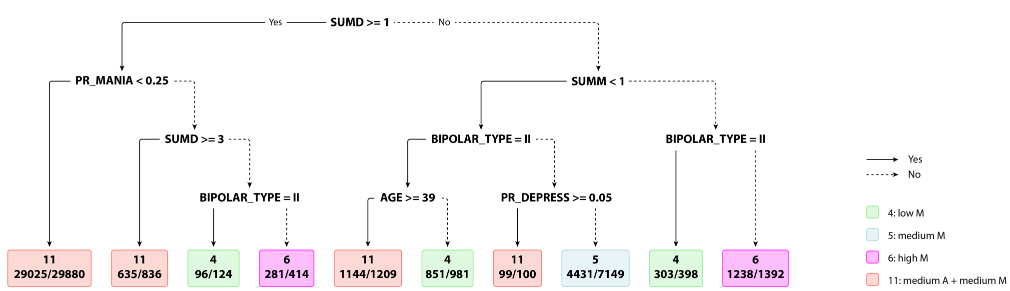

Figure C.1 Estimate optimal treatment regime by maximizing total discounted utility ($\gamma = 0.95$), which is projected onto a decision tree for interpretation. Each tree node shows the splitting criterion, majority treatment label and its proportion.