Hierarchical Hidden Markov Models for HIV-Transmission Behavior Outcomes

Li-Jung Liang
Department of Medicine Statistics Core
Email: liangl@ucla.edu
Joint work with Rob Weiss & Scott Comulada

Outline

- Motivation
  - Longitudinal Data Structure
  - Sexual Behavior and Substance Use
  - What and Why Hidden Markov Model?
- Our Approach
  - Bayesian Hierarchical Hidden Markov Models and computation
- Applications
  - Subsets of Previous Studies
- Preliminary Results
- Summary and Ongoing Work

2

Motivation

Longitudinal assessments of sexual behaviors and substance use

- Common in studies of cognitive behavioral intervention designed to reduce HIV-transmission behaviors
- Elicit frequencies of various sexual behaviors and of drug usage over a previous unit of time such as three months
- Study participants may not engage in transmission-associated behaviors at several time points
  - Truly abstinent or abstinent when assessed
  - Incarceration or in treatment
- Lead to zero-inflated data that is inadequately described by a unimodal count distribution
  - Distribution of non-zero counts is not well-described by a single Poisson distribution.
  - Commonly modeled using a ***Zero-Inflated Poisson or ZIP*** distribution.
  - For example, for recall of three month sexual behavior count, most non-zero observations will be small, typically in the low single digit values, but there will be several observations in the high 2 digits, or even in the 100s.

3

What's Hidden Markov Model?

- Hidden Markov models (HMM) describe the relationship between observed and underlying "hidden" (or unobserved) processes (Scott 2002 & 2005; Ridall et al 2005; Altman 2007, etc)
- The hidden process is assumed to follow a Markov chain, and the observed data are modeled as independent conditional on the sequences of hidden states.
- The observed data $\{Y_{ij}\}$ follow a HMM if
  1. The hidden states $\{S_{ij}\}$, $j=1,\ldots,n_i$, follow a Markov chain
  2. Given S_{ij}, Y_{ij} is independent of $Y_{i1},\ldots,Y_{ij-1},Y_{ij+1},\ldots,Y_{ni}$ and $S_{i1},\ldots,S_{ij-1},S_{ij+1},\ldots,S_{ini}$

4

Why Hidden Markov Model?

- Behavior and substance use data were mostly self-reported, which is generally subject to substantial variability due to measurement error.
- Consequences of this variability are that observed trajectories of the data values give a noisy representation of the true underlying evolving process.
- It is appropriate to treat the states as "unobserved" or hidden and to define a statistical framework that enables jointly characterizing the transitions between underlying hidden process and observed (noise) process.

5

Why Hidden Markov Model?

Have been used in many fields, such as genetic sequence alignment, gene profiling and recognition, carcinogenicity in animal studies, disease progression with longitudinal markers, speech recognition, computational biology, etc.

Pros
- Describe multi-model data better
- Source of variability can be appropriately identified
- Estimate the underlying behavioral patterns between different groups
- The separability of the model for the hidden process and the conditional model for the observed data leads to great flexibility in the overall model structure.

Cons
- Computation complexity, such as when the length of time-series observation is long
- Interpretation of hidden states becomes less clear when the number of states increases
- Lack of widely available software for fitting HMMs

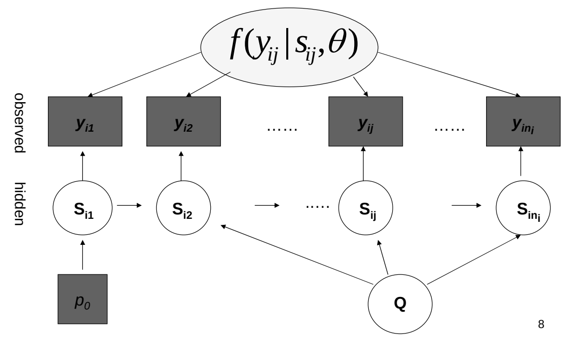
6

Applications

- Sexual behavior and substance use outcomes in studies of cognitive behavioral intervention
  - Number of sexual partners
  - Number of sex acts
  - Number of days of using hard drugs
- Long-term history of substance abuse research
  - Longitudinal usage of alcohol and illicit drugs data
  - Natural History Interview (NHI) Studies
  - Interested in learning “substance use career” or “pathway of substance usage”

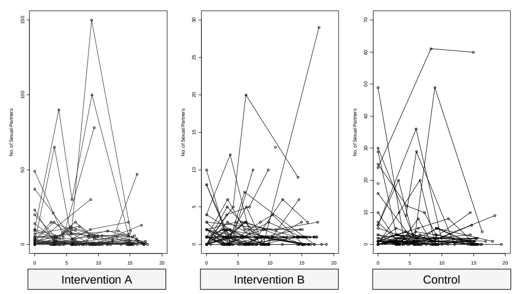
7

Data Structure



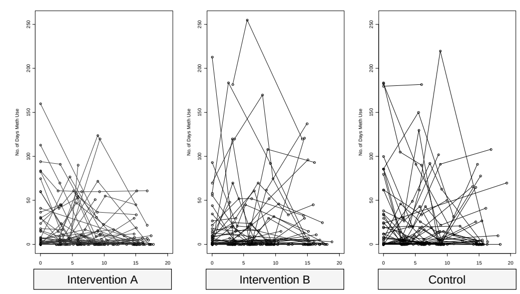
8

Number of Sexual Partners



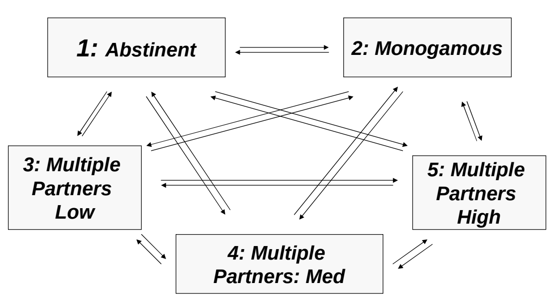
9

Number of Days - Hard Drug Usage



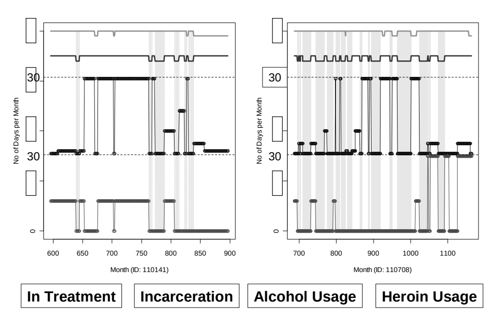
10

Hidden Markov States Sexual Behavior



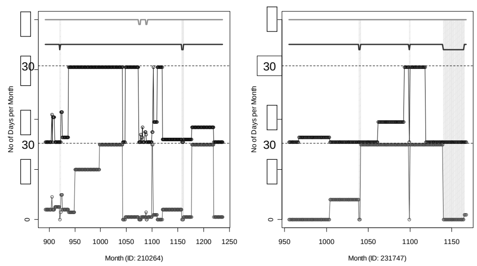
11

Heroin and Alcohol Usage



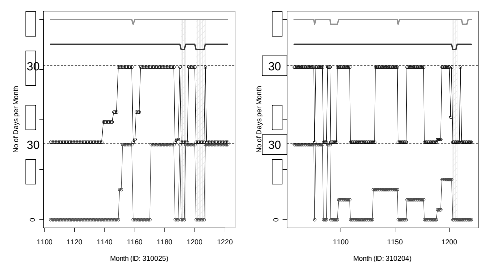
12

Cocaine and Alcohol Usage



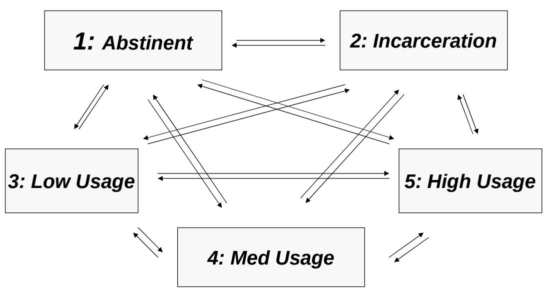
In Treatment | Incarceration | Alcohol Usage | Cocaine Usage

Meth and Alcohol Usage



In Treatment | Incarceration | Alcohol Usage | Meth Usage

Hidden Markov States Drug Usage



Hierarchical Hidden Markov Model

$$y_{ij}|s_{ij}\begin{cases}= 0, & \text{if } s_{ij}=1;\\ = 1, & \text{if } s_{ij}=2;\\ \sim \text{poi}(\lambda_{s_{ij}}) & \text{otherwise,}\end{cases}$$

where $s_{ij} \in \{3, 4, \ldots, S\}$, $\lambda_3 \le \lambda_4 \le \ldots \le \lambda_S$. Define $\theta = [\theta_3, \theta_4, \ldots, \theta_S]$ and $\theta_s = \lambda_s - \lambda_{s-1}$, where $s = 3, 4, \ldots, S$. Let Q be a SxS infinitesimal matrix, q_r is a Sx1 vector, $r = 1, \ldots, S$, and q_r' is the rth row of the Q matrix.

Model 1

Q involves covariates

Define $log(q_{i(-r)}) = X_{0i}\beta_0 + \ldots + X_{pi}\beta_P$, where $q_{i(-r)}$ is the rth row of the instantaneous transition rate matrix Q without the diagonal element q_{rr}.

$$\beta_p \sim \text{MVN}(\mu_p, \Sigma_p), \text{where } p = 0, \ldots, P$$

Random Effects Models

$$log(q_{i(-r)}) = X_{0i}\beta_0 + \ldots + X_{pi}\beta_P + Z_{ir}\eta_i + \varepsilon_{ir}$$

- **Model 2: Parametric Model**
  $\beta_p \sim \text{MVN}(\mu_p, \Sigma_p)$, where $p = 0, \ldots, P$
  $\eta_i \sim \text{MVN}(0, \Sigma_d)$
  $\varepsilon_{ir} \sim \text{MVN}(0, \Sigma_\varepsilon)$, where $\Sigma_\varepsilon = \sigma_\varepsilon^2 I$

- **Model 3: Non-Parametric Model**
  $\eta_i \sim G$, $G \sim \text{DP}\ (M, G_0)$,
  where $G_0 = \text{MVN}(0, \Sigma_d)$

Application
HIV-Transmission Behavior Outcome: Number of Sexual Partners

Study Description

- HIV-positive cohort (n = 175)
- Los Angeles, New York, San Francisco
- Entry criterion: substance use
- 16 –29 years old
- 26% Black, 42% Latino
- 69% gay men
- 3 groups: Control, Intervention-A (Int-A), and Intervention-B (Int-B)
- Followed for 15 months
- Goal: Reduce HIV transmission risk behaviors (risky sexual behavior & substance use)

Reference: Rotheram-Borus et al, 2004

Primary Outcomes

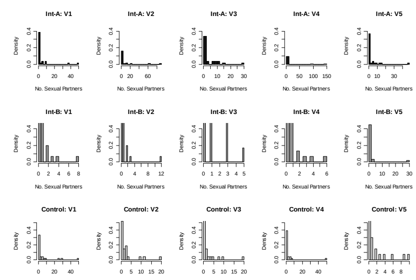
Risky sexual behaviors

- ***Number of sexual partners***
  - A sexual partner was defined as a male or female partner with whom the youth engaged in vaginal or anal sex.
  - Truly abstinent/monogamous or abstinent/monogamous when assessed?

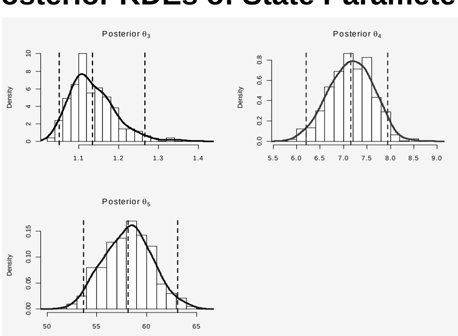
Substance use

- Observed frequency of past 3-month methamphetamine use
- High proportion of zeros
- Possibilities for zero counts: abstinent or abstinent when assessed

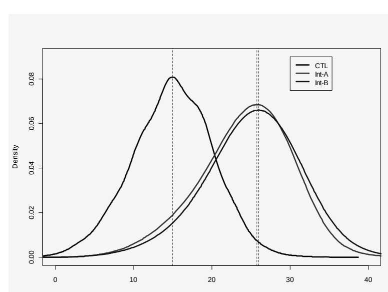
Number of Sexual Partners



Posterior KDEs of State Parameters



%Participants Stayed in Low Risk Group



Preliminary Results

- For illustration purposes, a subset of CLEAR data was used in this analysis.
- On average, about 25% of intervention participants stayed in the lower risk states (abstinent or monogamous) during the study, whereas less than 15% of the control participants stayed in the lower risk states (95% Credible Interval for the difference between the 2 groups: -2.6% - 22.2%) .
- In this subset, Int-B participants reported fewer number of sexual partners. We found that 84% of them (95% CI: 70% - 93%) stayed in the lower risk states (abstinent, monogamous, or low multiple partners) during the study.
- About 13% of the Int-A participants stayed in the higher risk states (95% CI: 9.7% - 16.1%) during the study.

25

Summary

- Bayesian hierarchical models are flexible to
  - handle the complex structure of data, for example, adequately describing multi-modal count distribution
  - HMM parameters estimated using Bayesian methods incorporate all sources of uncertainty, conditional on the model being correct (Scott et al 2005). Bayesian methods provide automatic measures of uncertainty even for complicated functions of the parameters.
- Advantages of our approach
  - Mixed-effects model approach, which is appropriate for making the individual inference and prediction
  - Incorporating covariates, such as treatment group, in the transition rate matrix
  - Allowing more states as compared to ZIP models and estimate the individual behavior pattern or substance usage pathway
- Difficulties of our approach
  - Computationally intensive
  - Interpretation of hidden states might be less clear when number of hidden states increases

26

Ongoing Work

- We demonstrated that the hierarchical hidden Markov models (CTHMM) are feasible for modeling longitudinal behavior outcomes (counts data).
- The preliminary findings based on the subset of previous studies are very interesting. There are other features of our approach that we want to explore. Finding ways of better summarizing and interpreting the results are also being investigated.
- We are currently working on
  - Improving computation algorithm for model 2,
  - extending models 2 to 3,
  - applying to the full data set (including the subjects with missing responses during the study), and other outcomes of interest (e.g., substance use).

27

Acknowledgement

UCLA Center for Community Health
Dr. Mary Jane Rotheram-Borus
Dr. Norweeta Milburn

UCLA Center for Advancing Longitudinal Drug Abuse Rsearch
Dr. Yih-Ing Hser

28